Bendik Lund

Host Genome Variation and Toxicity in Childhood Acute Lymphoblastic Leukaemia

Thesis for the degree of Philosophiae Doctor

Trondheim, February 2014

Norwegian University of Science and Technology
Faculty of Medicine
Department of Laboratory Medicine, Children's and Women's Health

NTNU – Trondheim
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Medicine
Department of Laboratory Medicine, Children's and Women's Health

ISBN 978-82-326-0016-8 (printed ver.)
ISBN 978-82-326-0017-5 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2014:41

Printed by NTNU-trykk

Genetisk variasjon og toksisitet hos barn med akutt lymfatisk leukemi

Akutt lymfatisk leukemi (ALL) er den vanligste kreftformen hos barn og i Norge får ca. 30-40 barn ALL hvert år. Behandlingen består av ulike cellegiftkurer gitt over en periode på 2,5 år og overlevelsen ligger i dag på ca. 80-85%. Behandlingen medfører mange bivirkninger, bl.a. nedsatt immunforsvar og infeksjoner. Av de som ikke overlever sykdommen skyldes ca. 75% kreftsykdommen i seg selv, mens 25% skyldes bivirkninger av behandlingen. Sistnevnte kalles behandlingsrelatert død (eng.: *treatment related death*, TRD).

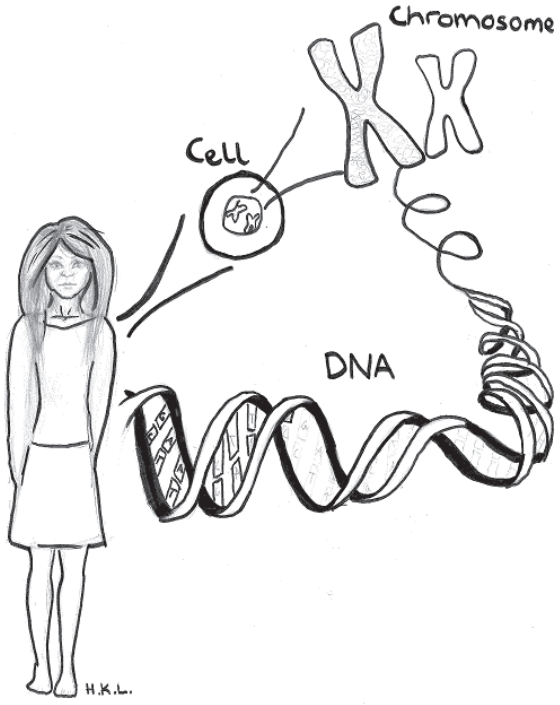
I denne studien har vi undersøkt forekomst, risikofaktorer og dødsårsaker hos 88 TRD-tilfeller blant 2700 pasienter med ALL i de fem nordiske landene Sverige, Danmark, Finland, Island og Norge. Vi fant at forekomsten av TRD var på 3,2% og at kjønn (pike), behandlingsrisikogruppe (høy-risk), T-celle sykdom, Down syndrom og gjennomgått beinmargstransplantasjon var risikofaktorer for TRD. Av dødsårsaker skyldtes 75% infeksjoner, 10% blødninger, 10% spesifikk organsvikt og 5% direkte kreftcelle-påvirkning (tumor byrde).

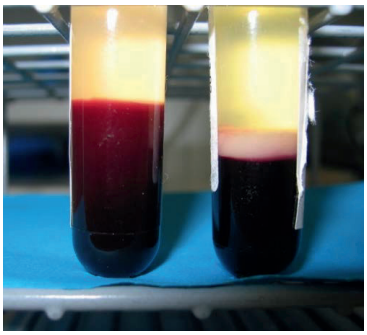
Vi har ingen god forklaring på hvorfor noen pasienter får flere og mer alvorlige infeksjoner enn andre. Det er grunn til å tro at genetiske faktorer spiller inn. Vi ville derfor undersøke om ulike mønstre av en vanlig forekommende variasjon i det menneskelige genom, nemlig enkeltnukleotidpolymorfisme (eng.: *single nucleotide polymorphism*, SNP) kunne spille en rolle for risiko for infeksjoner. Vi undersøkte 34000 SNP'er relatert til ALL hos barn og identifiserte en SNP-profil som med stor nøyaktighet kunne forutsi risiko for infeksjoner de første 50 dagene av behandlingen. Hvis disse funnene blir bekreftet i tilsvarende studier vil denne kunnskapen kunne brukes til å «skreddersy» behandlingen til enkeltpasienter og dermed redusere alvorlige bivirkninger.

I moderne genanalyser inngår ofte mange genvarianter samtidig. Det benyttes som regel DNA fra blodprøver, noe som kan være vanskelig å få tak i fra pasienter som ikke lenger er i live. Vi ville undersøke om det i stedet var mulig å gjøre genanalyser på arkivmateriale fra leukemipasienter, dvs. beinmargsutstryk og –biopsier. To typer markører ble undersøkt, 10 såkalte korte tandem repetisjoner (eng: *short tandem repeats*, STRs) og 34000 SNP'er. Resultatene ble sammenliknet med tilsvarende analyser på blodprøver og vi fant at 90% av STR-markørene lot seg gjenskape. For SNP-analysene fikk vi tilfredsstillende resultat for bare 7 av 34 prøver, men to av prøvene ga svært godt resultat. Dette illustrerer at bruk av gamle beinmargsprøver til genanalyser er mulig, men at metoden må forbedres.

Navn kandidat: Bendik Lund
Institutt: Institutt for laboratoriemedisin, barne- og kvinnesykdommer
Veiledere: Professor Helge Klungland, Professor Kjeld Schmiegelow, overlege Ann Åsberg, lektor/overlege Klaus Müller
Finansieringskilder: Samarbeidsorganet HMN-NTNU, Barnekreftforeningene i Danmark, Sverige og Norge; Forskningsrådet og Kræftens Bekæmpelse i Danmark; Otto Christensens Fond; The Novo Nordic Foundation; Barnekreftforskningsfondet, St. Olavs Hospital

Ovennevnte avhandling er funnet verdig til å forsvares offentlig
for graden PhD i molekylærmedisin.
Disputas finner sted i auditorium KA11, Kunnskapssenteret, St. Olavs Hospital/NTNU,
fredag 14. februar 2014, kl. 12.15.


Chromosome
Cell
DNA
H.K.L.



leukaemia [Gr. leukos, white, + haima, blood]. The tube to the left is a blood sample from a healthy person. The tube to the right shows a fresh taken blood sample from a boy newly diagnosed with acute lymphoblastic leukaemia with very high white blood cell count. The middle "milky" layer in the tube shows concentrated white blood cells (blast cells/cancer cells) illustrating the term 'white blood'.

(Photo by the author)

Contents

Acknowledgements/Collaborators

The present work has been carried out in the period 2009-2013 at the following locations/institutions: i) The Department of Laboratory Medicine, Children's and Women's Health, Faculty of Medicine, Norwegian University of Science and Technology, Trondheim, Norway; ii) Department of Pediatrics, St. Olavs Hospital, Trondheim, Norway; iii) Paediatric Oncology Research Laboratory, Bonkolab, Rigshospitalet, Copenhagen, Denmark; and iv) Center for Biological Sequence Analysis (CBS), Technical University of Denmark, Lyngby, Denmark.

This study has received financial support from The Liaison Committee between The Central Norway Regional Health Authority and The Norwegian University of Science and Technology (NTNU); The Danish Childhood Cancer Foundation; The Danish Medical Research Council for Health and Disease; The Danish Cancer Society; Otto Christensen's Foundation; The Novo Nordic Foundation; The Norwegian Childhood Cancer Organisation, Trøndelag County, Norway; Children's Cancer Foundation, Department of Pediatrics, St. Olavs Hospital; The Swedish Childhood Cancer Foundation.

For the completion of this thesis I am particularly grateful to my two main supervisors, Professor Helge Klungland and Professor Kjeld Schmiegelow whose enthusiasm and inspiration have followed me through this work. I will also express my gratitude to my two co-supervisors Ann Åsberg and Klaus Müller; Ann, because of her efforts in helping me initiate this project, and for her continuously support and numerous talks throughout the project; Klaus, because of his great supervision and specific knowledge within inflammation and immunity.

Additionally, this thesis had not been possible without great support from many others. My thanks go to:

- The patients and their families for taking part in this study and in NOPHO studies in general
- All co-authors in the separate studies who have been very supportive, and, not least, helping me transforming the manuscripts into "publishable" material
- Veslemøy who opened my eyes to laboratory work
- Louise and Agata at Bonkolab and CBS who, in addition to be co-authors, always were very supportive when talking and collaborating throughout the projects

(*cont.*)

- Kirsten who was the key laboratory "know-how" on library preparation and SNP sequencing
- Laeya for her participation as master student in our research group and sharing of common triumphs and failures in the laboratory
- The personnel at the pathology labs in Copenhagen and Trondheim
- Mats Heyman at the Childhood Cancer Research Unit, Stockholm, who has been of great help both as co-author and with data transmission from the NOPHO registry
- Jukka Kanerva, also co-author, who enthusiastically supported me as chair in the NOPHO Events Group
- All NOPHO workers who, through their daily efforts, provide data for the NOPHO registry
- The direction at the Department of Pediatrics, St. Olavs Hospital, and especially Dr Svein Kolmannskog (head of the Paediatric Oncology ward) who have been very interested and supportive to my project
- Espen and the rest of Journal Club for nice scientific discussions and sharing of common hang-ups of being a PhD-student
- My wife Line and my three beloved daughters Hanna, Sunniva and Vilde who have been so patient and encouraging with me during this work

Trondheim, 2nd November 2013

Bendik Lund

List of papers

Paper I

Lund B, Åsberg A, Heyman M, Kanerva J, Harila-Saari A, Hasle H, Söderhäll S, Jónsson ÓG, Lydersen S, Schmiegelow K; Nordic Society of Paediatric Haematology and Oncology. **Risk factors for treatment related mortality in childhood acute lymphoblastic leukaemia.** *Pediatr Blood Cancer.* 2011 Apr;56(4):551-9

Paper II

Lund B, Wesolowska-Andersen A, Lausen B, Borst L, Rasmussen KK, Müller K, Klungland H, Gupta R, Schmiegelow K. **Host genome variations and risk of infections during induction treatment in childhood acute lymphoblastic leukaemia.** Accepted for publication in *Eur J Haematol.*

Paper III

Lund B, Najmi LA, Wesolowska-Andersen A, Landsem VM, Rasmussen KK, Borst L, Gupta R, Schmiegelow K, Klungland H. **Archival Bone Marrow Samples: Suitable for Multiple Biomarker Analysis?** Accepted for publication in *Diagn Mol Pathol/Appl Immunohistochem Mol Morphol.*

Abbreviations

ALL	acute lymphoblastic leukaemia
ANC	absolute neutrophil count
CART	classification and regression tree
DNA	deoxyribonucleic acid
EFS	event-free survival
FFPE	formalin-fixed paraffin-embedded
FM	fluorometry
GC	glucocorticoids
HSCT	haematopoietic stem cell transplantation
MAF	minor allele frequency
MBL	mannose binding lectin
6-MP	6-mercaptopurine
NOPHO	Nordic Society of Paediatric Haematology and Oncology
OS	overall survival
SNP	single nucleotide polymorphism
SP	spectrophotometry
STR	short tandem repeats
TRD	treatment-related death
TPMT	thiopurine methyltransferase
WBC	white blood cell count
WGA	whole genome amplification

Definitions and key concepts

Acute lymphoblastic leukaemia (ALL): a cancer of white blood cells caused by genetic alterations of unknown origin leading to uncontrolled cell proliferation at a certain stage of maturation of the lymphoblasts.

Culture positive: the isolation of a microorganism from blood.

Gene library: a stored collection of DNA fragments ready for further processing (sequencing); in our study, the library was prepared from the patient's genomic DNA and includes the SNP-containing DNA fragments.

Genetics: the science of genes, heredity and variation in living organisms.

Genomics: a discipline of genetics that applies DNA sequencing methods and bioinformatics to sequence and analyse the function and structure of the genome (the complete set of DNA within a single cell of an organism).

"Infectious event": the combination of fever or other signs of infection and initiation of antimicrobial therapy after the initiation of anti-leukaemic therapy.

Metabolic pathway: a series of chemical reactions within a cell often catalysed by enzymes.

Neutropenia: absolute neutrophil count of $\leq 0.5 \times 10^9$/L.

Remission: a state in which the vast majority of malignant blast cells have been removed from the patient by induction treatment, usually when the number of malignant blast cells in the bone marrow is $<10^{-3}$ cells/μL.

Remission induction treatment: the period of treatment (with chemotherapy, glucocorticoids and supportive care) from the point of diagnosis to the time that remission is achieved.

Single nucleotide polymorphism: a phenomenon present at various frequencies throughout the DNA molecule involving variation in a single nucleotide (A, T, C or G).

Sammendrag på norsk

Kreft er den vanligste medisinske dødsårsaken hos barn mellom 1 og 15 år. Akutt lymfatisk leukemi (ALL) er den vanligste kreftformen hos barn og i Norge får ca. 30-40 barn ALL hvert år. Behandlingen strekker seg over 2,5 år og består av 6-12 ulike cellegifter i forskjellige kombinasjoner i tillegg til støttebehandling og overvåkning. Behandlingen medfører mange bivirkninger, bl.a. nedsatt immunforsvar som gir økt risiko for infeksjoner. Den totale overlevelsen for barn med ALL er i dag på ca. 80-85%. Av de som dør etter å ha fått diagnosen skyldes dette for ca. 75% kreftsykdommen i seg selv hvorav de fleste skjer etter tilbakefall av sykdommen (residiv-relaterte dødsfall). De resterende 25% av dødsfallene skyldes bivirkninger av behandlingen, noe som kalles *behandlingsrelatert død* (TRD, *treatment-related death*).

I denne studien har vi undersøkt forekomst, risikofaktorer og dødsårsaker hos 88 TRD-tilfeller blant 2700 pasienter med ALL i de fem nordiske landene Sverige, Danmark, Finland, Island og Norge. Forekomsten av TRD var 3,2% og denne holdt seg uendret fra den første ALL protokollen (NOPHO ALL 1992) til den neste (NOPHO ALL 2000). Risikofaktorer var kjønn (pike), behandlings-risikogruppe (høy-risk), T-celle sykdom, Down syndrom og gjennomgått beinmargstransplantasjon. Av dødsårsaker skyldtes 75% infeksjoner, 10% blødninger, 10% spesifikk organsvikt og 5% direkte kreftcelle-påvirkning (tumor byrde).

Vi har ingen god forklaring på hvorfor noen blir mer syke av behandlingen enn andre. Det er derfor grunn til å tro at genetiske faktorer spiller inn. I de siste 10 årene har man forstått vesentlig mer om den naturlige genetiske variasjon hos menneske. Vi ville undersøke om ulike mønstre av en vanlig forekommende variasjon i det menneskelige genom, nemlig enkeltnukleotidpolymorfisme (eng.: *single nuclotide polymorphism*, SNP) kunne spille en rolle for risiko for alvorlige infeksjoner under leukemibehandlingen. Vi undersøkte 34000 SNP'er relatert til ALL sykdommen hos barn og fant 24 SNP'er som var assosiert til risiko for infeksjon under behandlingen. Ved hjelp av CART-analyse (*classification and regression tree*) identifiserte vi 4 SNP'er som inngikk i en SNP-profil som var i stand til å forutsi risiko for infeksjoner de første 50 dagene av behandlingen med stor nøyaktighet. SNP'ene tilhører genene *OR51F1, CBR1, POLDIP3* og *CCL11* som bl.a. regulerer cellegiftomsetning, cellevekst og betennelsesreaksjoner. Hvis disse funnene blir bekreftet i tilsvarende studier vil denne kunnskapen kunne brukes til bedre å «skreddersy» behandlingen til enkeltpasienter og dermed redusere forekomsten av alvorlige bivirkninger.

I moderne genanalyser inngår ofte mange genvarianter samtidig. Det benyttes som regel DNA fra blodprøver, noe som kan være vanskelig å få tak i fra pasienter som ikke lenger er i live. Vi ønsket derfor å undersøke om det var mulig å analysere biologiske markører på arkivmateriale fra leukemipasienter, dvs. beinmargsutstryk og -biopsier. DNA ble ekstrahert fra beinmargsprøvene og deretter kvantitert. Pga. små mengder DNA fra hver pasientprøve ble det gjort hel-genom amplifisering (*whole genome amplification,* WGA) som et forsøk på å øke DNA mengden i prøvene. Deretter ble prøvene analysert for to typer gen-markører, 10 såkalte korte tandem repetisjoner (eng.: *short tandem repeats*, STRs) og 34000 SNP'er. Resultatene ble sammenliknet med tilsvarende markørundersøkelser på blod og vi fant at 90% av STR-markørene lot seg gjenskape fra arkivmaterialet. For SNP-markørene fikk vi tilfredsstillende resultat for bare 7 av 34 prøver, men to av prøvene ga svært godt resultat. WGA fungerte ganske bra for noen av prøvene, men dårligere sammenliknet med prøver uten WGA. Konklusjonen ble at bruk av gamle beinmargsprøver til multiple genanalyser er mulig, men analysemetoden må forbedres.

Summary in English

Cancer remains the leading cause of disease-related mortality among children aged 1-14 years in developed countries. Acute lymphoblastic leukaemia (ALL) is the most common malignancy in childhood and accounts for about 25% of all childhood cancers. In Norway about 30-40 children are diagnosed with ALL each year. Treatment comprises of a combination of 6-12 different chemotherapy drugs administered over a period of 2.5 years with additional supportive care. Side effects remain a great challenge for both patients and physicians. One of the most common side effects is immunosuppression which increases the risk of infection.

Today, the overall survival rate in childhood ALL is 80-85%. About 75% of deaths related to ALL are caused by the disease itself, most after relapse, but about 25% of deaths are caused by treatment and are classified as *treatment-related deaths* (TRDs).

In this study, the incidence, risk factors and causes of death were investigated for 88 TRDs among 2,700 patients in the Nordic countries Sweden, Denmark, Finland, Iceland and Norway. The incidence of TRD was 3.2%, which was stable in the two protocols, the Nordic Society of Paediatric Haematology and Oncology (NOPHO) ALL 1992 and ALL 2000

protocols. The risk factors identified were female gender, treatment risk-group, T-cell disease, Down syndrome and stem cell transplantation. Seventy-five percent of the deaths were related to infection, 10% to bleeding or thrombosis, 10% to organ failure and 5% to the tumour burden.

We have no good explanation as to why some patients experience more severe infections than others, even when receiving the same drugs, drug dosage and supportive care. It is reasonable to believe that genetic factors play a role. In accordance with the increasing understanding of host genetic variation the past 10 years, we investigated 34,000 single nucleotide polymorphisms (SNPs) relevant in childhood ALL in a Danish ALL cohort including 69 children and detected 24 SNPs which associated with infections during induction treatment. Through CART-analysis (*classification and regression tree*) a four-SNP risk profile were identified as highly predictive of risk of infections during the 50 days induction treatment. The four SNPs belong to the genes *R51F1, CBR1, POLDIP3* and *CCL11* which regulate drug metabolism, cell-growth and inflammation. If these findings are replicated in larger studies, such knowledge may be useful for developing personalized medicine with more tailored therapy and supportive care, which will hopefully reduce the severity of side effects and increase overall survival.

Studies involving multiple biological markers, such as SNPs and short tandem repeats (STRs), often use high-quality DNA extracted from blood samples. However, blood samples are not always easy to obtain from study participants, especially if the patient has died. Archival bone marrow samples from patients with leukaemia are available and represent a potential DNA source. We were interested in exploring whether such archival material is usable for the analysis and identification of multiple markers. DNA from 21 bone marrow smears and 13 bone marrow biopsies were extracted and quantified. Because of small amounts of DNA from each sample whole genome amplification (WGA) was applied. Ten different STR-markers and 34,000 SNPs were analysed and compared with corresponding blood samples. For the STR markers 90% detection rates were obtained. Shorter markers (107bp-242bp) from samples stored 0-3 years gave better results compared with longer markers (219bp-317bp) stored 4-10 years. Multiple SNP analysis was more complicated and only seven of 34 archival samples gave acceptable results (SNP call-rates above 50%). However, in two samples, nearly 100% of SNP-markers were detected. Although increasing the total amount of DNA, WGA reduced the analysis quality. In conclusion, DNA from archival bone-marrow samples might be used in multiple marker analysis, but adjustments of the laboratory set-up are essential to optimize this method.

INTRODUCTION

Topic of the thesis

This thesis examines treatment-related death (TRD) and treatment-related infections in Nordic patients with childhood acute lymphoblastic leukaemia (ALL). Clinical risk factors, host genome variation, metabolic pathways, and use of archival samples for DNA extraction and multiple biomarker analysis are explored.

Perspective and rationale

Side effects from cancer treatment are daily challenges for patients and clinicians. Severe side effects such as infection, occasionally leading to TRDs, represent the most tragic outcome of therapy-induced toxicity, especially for childhood ALL treatment. Increased understanding of risk factors, included the possible contributions of host genetic factors, is the first step towards decreasing the occurrence of such severe events. The role of host genomic variation and treatment-related side effects has not been explored fully for childhood ALL. In our modern high-throughput sequencing era, the use of different DNA-sources, such as blood samples, tissues and smears, has great potential in studies of host genomic variation. Blood samples can be difficult to collect in some situations, for instance from patients who have died. Alternative samples from such patients are archival samples collected earlier as part of the hospital's routine. For leukaemia patients, bone marrow smears and biopsies represent a potential DNA source for such studies.

We conducted this study on TRDs and infections during treatment by integrating clinical and genetic information. We aimed to understand more deeply the causes and risks of treatment-related infections and TRDs. Bone marrow archival samples were also tested to determine the suitability of DNA extraction and marker-analysis.

The study is organized in three parts. **Study I** was a clinical study of the risk factors and causes of death in 88 TRDs among 2,735 Nordic childhood ALL patients treated between 1992 and 2008; the data were from the Nordic Society of Paediatric Haematology and Oncology (NOPHO) registry. **Study II** combined clinical data on infections during ALL induction treatment with multiple SNP data including 34,000 SNPs aiming at identifying genetic risk profiles and relevant metabolic pathways related to infections during treatment. **Study III** was a methodological study exploring the challenges involved in using DNA derived from archival bone marrow samples from childhood ALL patients in multiple

biomarker analysis. These studies are referred to as the paper produced from the data in the study: Papers I, II and III, respectively.

Theory and central concepts

Acute lymphoblastic leukaemia

Cancer remains the leading cause of disease-related mortality in developed countries among children aged 1-14 years (Ghaderi et al., 2012; "Statistics Norway [Internet] Available from: https://www.ssb.no,"). ALL is the most common malignancy in childhood, accounting for about 25% of all childhood cancers (Kaatsch, 2010). ALL is a cancer of the white blood cells (WBC) of unknown origin. The disease develops as a consequence of malignant transformation of a single abnormal lymphoid precursor cell leading to uncontrolled proliferation and expansion in the bone marrow and other lymphoid organs. In the 1950s, childhood ALL was considered an incurable disease. During the past few decades, there has been a steady increase in both event-free survival (EFS) and overall survival (OS); 75% of all patients are now cured by first-line therapy, and there is a long-term survival rate of about 83% (Fig. 1) (Schmiegelow et al., 2010).

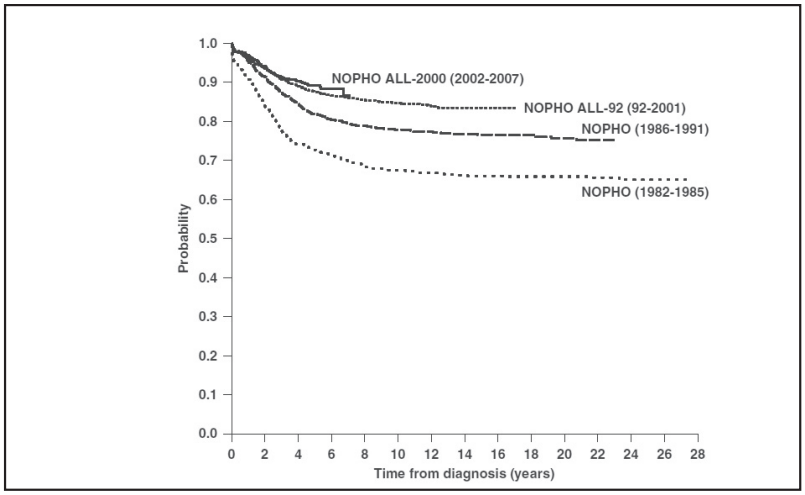


Figure 1. Overall survival in four consecutive Nordic Society of Paediatric Haematology and Oncology (NOPHO) cohort periods (4). Number of patients in the respective protocols, NOPHO1982-1985: n_1=719; 1986-1991: n_2=937; 1992-2001: n_3=1,645; and 2002-2007: n_4=1,023

ALL is diagnosed by the presence of more than 25% lymphoblast cells in the bone marrow recognized by microscopic examination of Giemsa-stained bone marrow smears and flow cytometric techniques. Treatment comprises metronomic administration of 6-12 different chemotherapeutic agents for 2.5 years from the time of diagnosis. For patients with very high relapse risk, haematopoietic stem cell transplantation (HSCT) is the only treatment option providing a chance of a cure. Chemotherapy treatment implies side effects, and intensive treatment requires intensive supportive care. Treatment-related toxicity is a great challenge for both patients and clinicians, and TRD is the most severe side effect (Blanco et al., 2012; Prucker et al., 2009; Rahiala et al., 1998). Short- and long-term side effects play a crucial role in determining the patient's quality of life both during and after treatment. Examples of side effects are infections, bleeding, thrombosis, cardiovascular and pulmonary problems, neurological impairment, nutritional problems, osteonecrosis, fatigue and secondary malignancies. In addition, the risk of relapse is always present during treatment and the years following treatment.

The events contributing to reduced EFS are resistant disease (1.0%), relapse (16.9%), TRD (3.2%), and secondary malignant neoplasms (SMNs) (1%). Factors influencing these events are: i) the malignant lymphoblastic clone with its intrinsic genetic and chromosomal rearrangements; ii) anti-leukaemic treatment including chemotherapy and glucocorticoids (GCs); iii) supportive care (e.g., monitoring/nursing, fluids, nutrition, antibiotics, anti-thrombotics and procedures); and iv) host factors (e.g., host genetics and host environment). The most crucial treatment period is the first weeks and months after diagnosis because the malignant clone is still present (first days/weeks) and may infiltrate into the bone marrow, leading to neutropenia, lymphopenia, anaemia and thrombocytopenia, which increases the risk of bleeding and infections. The malignant cells can also infiltrate into other vital organs, thus increasing the risk of leukostasis. In addition, initiation of chemotherapy and GCs (prednisolone/dexamethasone) leads to immunosuppression, risk of bleeding and problems related to tumour-lysis (e.g., renal failure).

Treatment-related death

Despite improved supportive care, 2-4 % of patients die because of treatment-related causes. The risk of TRD has not decreased substantially in the past 20 years (Blanco et al., 2012; M. S. Christensen et al., 2005; Hill et al., 2004; Prucker et al., 2009; Rubnitz et al., 2004) and TRDs contributes to reduced outcomes of cancer treatment in general. Normally, a TRD is

regarded as a death during treatment, and in some studies, the term also includes death occurring a short time after the end of treatment. Death many years after the end of treatment that is related to late effects after cancer treatment is not normally regarded as a TRD.

In parallel with the decreasing number of patients dying from the leukaemia itself (i.e., tumour burden-related death, resistant disease or relapse), TRDs seem to comprise an increasing proportion of overall mortality because the TRD rate seems to be unchanged (Blanco et al., 2012). In both the NOPHO ALL-1992 and ALL -2000 protocols, TRDs accounted for 25% of all deaths (Fig. 2). One possible explanation for the stable TRD-rate is that the reduction in relapse rates by increased treatment intensity for high-risk patients has led to a higher risk of TRD. An increased focus on supportive care (i.e., more adequate antibiotics, better treatment of tumour-lysis and more awareness of side effects in general), may have balanced this increased toxicity, thus producing stable TRD rates. Thus, as the proportion of deaths attributed to TRDs has increased, the focus on severe toxic events has become increasingly important.

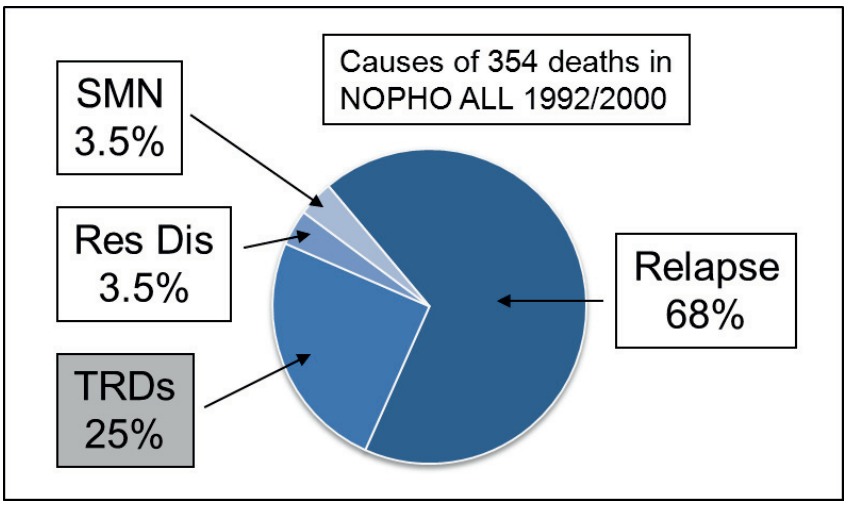


Figure 2. Causes of death in NOPHO ALL-1992 and ALL-2000 protocols**.** SMN, second malignant neoplasm; Res Dis, resistant disease; TRDs, treatment-related deaths (data from NOPHO ALL registry)

Definition of TRD

There is no internationally accepted definition of TRD (also referred to as *toxic death, treatment-related mortality* or *non-relapse mortality*). In clinical trials, a death during therapy is reported as a *serious adverse event*, which is defined as any undesirable experience associated with the use of a medical product in a patient, and includes (among others) death, life-threatening event, hospitalization (initial or prolonged), disability or permanent damage. This term also includes non-fatal events and is mostly used in studies of new drugs (Choueiri et al., 2013; "U.S. Food and Drug Administration [Internet] Available from: http://www.fda.gov,"). Another term is *suspected unexpected severe adverse reaction*. Within the field of paediatric oncology, most authors include as TRDs those events related to a certain drug or drug combinations, as well as those related to the total therapy given, namely chemotherapy, supportive care and procedures (HSCT, surgery, cardiopulmonary support, etc.).

Many authors define TRD differently, as reported by Ethier and co-workers (Ethier et al., 2011). Some authors relate TRD to the phase of therapy during which the event occurred (e.g., induction, consolidation or maintenance treatment). Others relate TRD to specific time sets after the start of therapy. TRDs after HSCT, early deaths (i.e., after diagnosis, but before treatment) and deaths happening soon after the end of treatment (but obviously related to treatment) are sometimes included, but sometimes not. Authors do not necessarily present the overall toxicity-related death rate over the course of treatment when presenting protocol results and, because of the variability in the definition, TRD-rates can be difficult to interpret and may not be comparable between studies (Ethier et al., 2011).

Causes of TRDs - attribution

Another problem with reporting TRDs is attribution or the cause of death. An internationally acceptable definition or a system for attribution does not exist. Authors use different definitions of infection-related mortality (M. S. Christensen et al., 2005; Slats et al., 2005). There is no definition of the type or extension of haemorrhage that constitutes a bleeding related death. Patients might have multiple concurrent serious events close to death (e.g., organ dysfunction, infection and haemorrhage), complicating the identification of the primary cause of death. Ethier and co-workers (Ethier et al., 2011) proposed the following:

- A system is needed to categorize the certainty of attribution as definite, probable or possible.
- In cases with multiple causes of death, all possible causes of death along with the associated certainty of attribution should be listed.

These proposed standards would make it easier to compare results between treatment protocols.

In Paper I, TRD-patients were grouped by cause of death into five categories as follows: i) tumour burden-related, ii) bleeding or thrombosis, iii) infection, iv) organ toxicity and v) other or uncertain cause. This grouping was an attempt to increase our understanding of the cause of death, even though we realize that such grouping is complicated. However, our study collected detailed clinical information on the causes of death (Appendix), including autopsy reports for some patients, and we believe this information is meaningful. An additional problem is a death by multiple causes. When in doubt, cases included in our study were discussed within the NOPHO Events Group as an attempt to clarify the probable chain of events leading to death, with special emphasis on the most important initiating factor. An example is a patient with *Pseudomonas*-sepsis who suddenly deteriorates, is put on a ventilator and, secondary to the infection, exhibits severe coagulation disturbances which lead to a fatal brain haemorrhage. In this case, we considered the *Pseudomonas* infection as the most important initial event and coagulation disturbances as the contributing cause. This case was classified as an infectious TRD.

Infections and inflammation during treatment

Infections are one of the most common side effects of treatment (Afzal et al., 2009; Rahiala et al., 1998; Rungoe et al., 2010). Infections can be life-threatening and are the most common cause of TRD (Study I) (M. S. Christensen et al., 2005; Prucker et al., 2009). The clinical spectrum of symptoms and signs of infection/inflammation during treatment spans from mild symptoms of low-grade fever but otherwise no symptoms to severe life-threatening infections requiring cardiopulmonary support. A widely used clinical entity is the condition termed "febrile neutropenia" which happens to all patients several times during treatment (Stabell et al., 2008). This occurs when the patient displays neutropenia (i.e., absolute neutrophil count (ANC) $<0.5 \times 10^9$/L) and the body temperature is >38.0-38.5 °C.

When a patient experiences febrile neutropenia, most treatment centres start routinely with empirical broad-spectrum antibiotics. The patient is monitored closely and, if fever disappears and no other signs develop, antibiotics are terminated after 3-5 days or are given until the patient is no longer neutropenic and is otherwise in good condition. In other cases, the initial mild symptoms develop into a more severe episode with fulminant sepsis. Sometimes febrile neutropenia is caused by an infection. In other situations, febrile neutropenia probably represents a state of inflammation without infection, giving rise to fever.

A documented infectious focus and/or a microbiological organism is detected in 20-40% of patients with a suspected infectious episode in studies of childhood leukaemia (Bakhshi et al., 2008; Graubner et al., 2008; Lehrnbecher et al., 2004). The type of infective organism ranges from common Gram-positive and Gram-negative bacteria to microorganisms found more often in immunosuppressed patients (e.g., *Pseudomonas* species, fungi, *Pneumocystis jiroveci*). Sometimes these episodes lead to death (infectious TRD). From a clinical point of view, it cannot be predicted which patients will experience the most severe infectious episodes because patients within the same risk group (i.e., same treatment intensity) and of the same age/gender may have a completely different "infectious profile" throughout treatment. Some patients will develop mild and few infectious episodes, whereas others will develop more severe infectious episodes. This clinical observation leads to the idea that, in addition to known risk factors and incidental differences between individuals, the host genetic profile might contribute to greater phenotypic variation. Childhood ALL patients are immunocompromised by the disease and chemotherapy given. Additionally comes, for young children, an immature immune system. Thus, the host genomic profile relating to innate immunity and pharmacogenetics might play a more important role in the response to an infectious episode in ALL patients compared with non-immunocompromised and otherwise healthy individuals.

Infections are complex diseases in which the outcomes depend on the interaction between the host's genetic make-up and the environment. Important environmental factors are i) the type of infectious microorganism, ii) antibiotic susceptibility, iii) the site of infection, iv) how soon the infection is detected, v) appropriate antibiotic treatment and vi) supportive care (resuscitation, supportive medical management and/or surgery) (Huttunen et al., 2011; Sutherland et al., 2009). The effect of the host's genetic make-up may also be modified by environmental factors.

The pathogenesis of sepsis includes the involvement of inflammatory cells and signalling molecules, and the balance between the pro- and anti-inflammatory states. This

balance might be disturbed, leading to an uncontrolled pro-inflammatory state with a potentially fatal outcome. This is also reflected by the complexity at the genomic level, in which host defence against an invasive microorganism involves the interplay of hundreds of signalling molecules and increased expression of multiple genes (Calvano et al., 2005; Huttunen et al., 2011). Factors that might contribute to the outcome in infectious events during ALL treatment are shown in Fig.3.

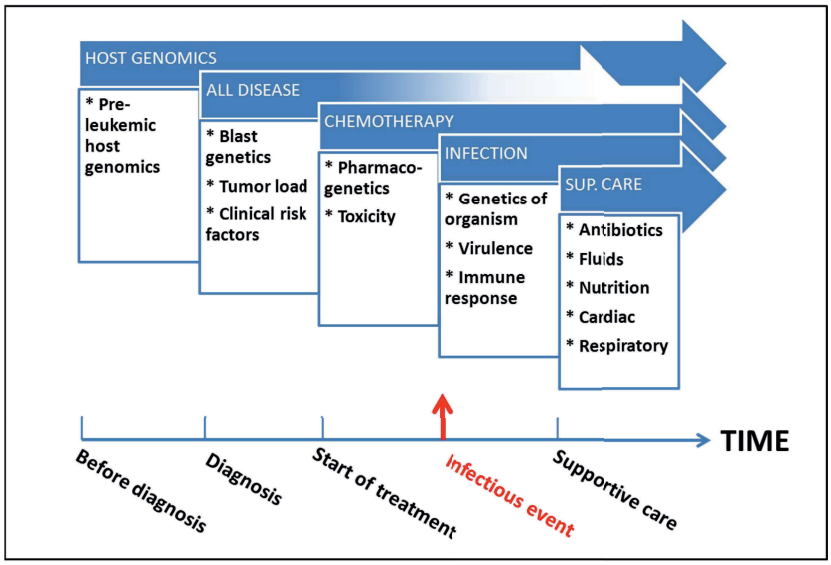


Figure 3: A model illustrating possible factors influencing the outcome of infectious events during ALL treatment (from Paper II).

Host genome variation

In genetic studies of leukaemia, research has focused on genetic alterations in leukaemic blast cells to understand the disease aetiology and pathogenesis (Mullighan, 2012). Less attention has focused on host genomics and its influence on outcomes. Host DNA sequence variation influences many aspects of a person's characteristics including anthropometry, risk of disease, drug metabolism and other environmental responses (Sachidanandam et al., 2001). Examples of host genome variation are SNPs, copy-number variation, repeating elements (microsatellites, short tandem repeats (STRs), minisatellites) and inversions. A SNP is a single nucleotide variation present at various positions throughout the DNA molecule. This phenomenon occurs when a single nucleobase (A,T,C or G) at a specific position in the DNA strand is substituted with another nucleobase. In Fig. 4, thymine (T) in Allele 1 is substituted with cytosine (C) in Allele 2. The most prevalent allele in the population is termed the wild-type allele, and the less prevalent allele is termed the variant allele. Individuals can have two copies of either the wild-type or the variant allele, called homozygous, or can have one copy each of the wild-type and variant allele, called heterozygous. Most SNPs are bi-allelic. SNPs occur at certain frequencies in different populations and can be assigned a minor allele frequency (MAF) in a population, i.e., the frequency of the variant allele.

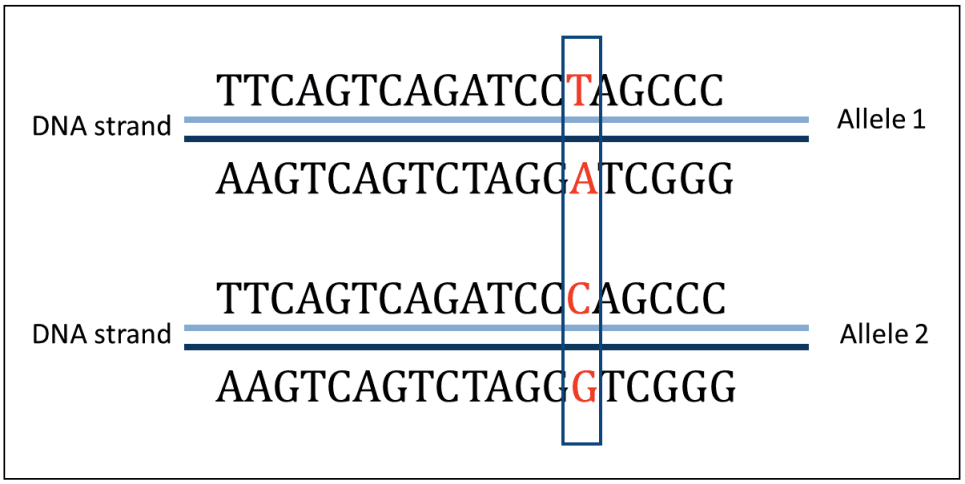


Figure 4. Single nucleotide polymorphism. In Allele 2, a T is substituted with a C in the corresponding position as in Allele 1. The nucleobases in double-stranded DNA are complementary so that T (thymine) always binds to A (adenine) and C (cytosine) always binds to G (guanine).

SNPs are highly frequent. About one SNP occurs per 300-1,000 base pairs (K. Christensen et al., 2007; Sachidanandam et al., 2001), meaning that up to 10 million SNPs are situated in the 3 billion base-pairs in the entire human genome. SNPs can be located in exons, introns or close (promoter) or remote regulatory (enhancer) regions in the DNA. In exons, they can be synonymous (i.e., they do not affect the amino-acid sequence in the protein) or non-synonymous. Non-synonymous SNPs give rise to changes in the amino-acid composition of the corresponding protein. Since the first publication of the entire sequence of the human genome in 2001 (Consortium, 2004; Sachidanandam et al., 2001), much attention has focused on linking phenotypical traits (e.g., diseases, disease risk factors, enzymes, drug metabolism) with human genetic variants including SNPs. It is important to recognize that the association of one or more SNPs with a certain trait or phenotype does not necessarily explain the causality of the underlying biological mechanism. The associated SNP can be causal, but it can also be a "tag-SNP" in linkage disequilibrium with the causative SNP because gene fragments are inherited as haplotypes.

An example of a known clinically relevant SNP used in treatment of childhood ALL is genetic variation in the *TPMT*-gene. 6-Mercaptopurine (6-MP) has been used for many years in the treatment of childhood ALL and is one of the main drugs used during maintenance therapy in most protocols. 6-MP is metabolized by thiopurine methyltransferase (TPMT), an enzyme encoded by the *TPMT*-gene. This enzyme is affected by functional SNPs in the *TPMT*-gene that reduce the activity of the enzyme (Schmiegelow et al., 2009). A consequence for heterozygous or homozygous (variant allele) patients is that reduced 6-MP metabolism because of reduced TPMT activity can increase drug toxicity and cause prolonged neutropenia, which increases the risk of infections. Consequently, the dose of 6-MP is reduced for such patients. *TPMT*-SNP-genotyping is performed routinely in many contemporary ALL protocols including the NOPHO ALL-2008 protocol.

STRs or microsatellites are small DNA elements comprising one-, two-, or three-base sequences repeated in tandem 15-100 times. STR loci often mutate into multiple alleles and are highly polymorphic in the number of repeats. They are stable, highly polymorphic DNA markers often used in family studies, in forensic medicine and in paternity testing. In our studies both multiple SNPs and STRs were included as biological markers.

Multifactorial traits such as infection are considered to be affected by multiple underlying causal pathways (Huttunen et al., 2011; Sutherland et al., 2009). These pathways include gene regulation from single genes to protein formation through transcription, translation, protein stability, RNA interference, DNA methylation, and other cellular events.

Based on this complex nature of several important disease traits, including the multifaceted interplay of genes and proteins, multiple genetic studies often focus on pathways, gene-networks and gene-gene interactions.

Genome-wide association studies (GWASs) have been useful in many studies (Andersson et al., 2009; Yang et al., 2009). Using this technique, one can study several SNPs at a time (0.5-1 million) in a single run, and newly identified genes or SNPs can be linked to a phenotype. A common limitation of GWAS design is the microchip-based design with a fixed SNP set-up and low flexibility. Additionally, most SNPs are located outside known genes. If a specific SNP is associated with a certain trait, interpreting the underlying biological mechanism is complicated. Another approach is single-SNP or single-candidate gene-association studies. The benefit of such studies is that the function of the candidate gene is known, but the obvious limitation is identifying the pathway to be studied.

Genetic variation and infections in childhood acute lymphoblastic leukaemia

Host genomic variation and risk of infections have been investigated in many groups of patients, including children and childhood cancer patients. One important question is whether immunocompromised childhood ALL patients have the same genetic risk profile regarding severity of an infectious episode as non-immunocompromised patients. The outcomes of infection depend on many factors to which the genetic contribution probably varies. The infective microorganism's susceptibility to the antibiotics used probably contributes more to the risk of severe outcome compared with the involvement of a specific SNP; for instance, in innate immunity in an otherwise healthy child. However, many genetic polymorphisms that are significantly associated with more severe outcomes of infection have been found. Selected studies, including some studies on children with leukaemia, are listed in Table I. The studies in Table I show only a few SNPs or other genetic variations associated with infection. Few studies have investigated multiple biomarkers related to infection or the metabolic pathways of importance for determining the outcome of infection.

Gene/protein	Disease	Population	Year	Ref.
Complement factor H Y402H, MBL2, PAI-1, CFB, C1qA	SIRS, sepsis	**Children**	2010	(Agbeko et al., 2010)
TNF-α, LT-α	Lethal infection, **ALL**	**Children**	2009	(Kidas et al., 2009)
IL7Rα	Outcome HSCT	**Children**/ adults	2006	(Shamim et al., 2006)
TLR2/4/5, CD14, MBL, IRAK4, TNF-α, MIF, IL-10, IL-6, IL-1RA	Sepsis (many microorganisms)	**Children**/ adults	2005	(Arcaroli et al., 2005) (review)
IL6, CHIT	G-negative bacteria, AML	**Children**	2005	(Lehrnbecher et al., 2005)
MBL	Infections, **ALL**	**Children**	2005	(Lausen et al., 2006)
IL10 promoter (TNF, LT-α)	*Pneumococcus*, septic shock	adults	2003	(Schaaf et al., 2003)
IL-12, IL-12R, IFN-γ, IFN-γR, STAT1	Mycobacteria, *Salmonella*	**Children**/ adults	2002	(Ottenhoff et al., 2002) (review)
TLR4	G-negative infections	Adults	2002	(Agnese et al., 2002)
TNF2 (TNFα-promoter)	Septic shock, mortality	Adults	1999	(Mira et al., 1999)

Table I: Selected reports on genetic polymorphisms affecting the outcomes of infections for children and adult cancer and non-cancer patients. MBL, mannose-binding lectin; PAI-1, plasminogen activator inhibitor-1; CFB, complement factor B; C1QA, complement C1q subcomponent subunit A; TNF, tumour necrosis factor; LT, leukotriene; IL, interleukin; TLR, toll-like receptor; CD, cluster of differentiation; IRAK, interleukin receptor-associated kinase; MIF, macrophage migration inhibitory factor; CHIT, chitotriosidase; AML, acute lymphoblastic leukaemia; STAT, signal transducer and activator of transcription; SIRS, systemic inflammatory response syndrome; HSCT, haematopoietic stem cell transplantation.

Biological samples for genetic studies

Multiple genetic association studies normally depend on high-quality DNA stored in biobanks. Considering the clinical risk factors for TRD identified in Study I, we were interested in including host genomic variations of possible relevance to infectious TRD in our studies. Unfortunately, there was no systematic biobanking of blood samples from the TRD patients, and the question was raised whether other DNA-sources could be used for further genetic studies. Archival samples, such as biopsy specimens, cytology and haematology smears, have been used in studies of genetic biomarkers for decades (Boyle et al., 1998; Kanteti et al., 2009; Lausten-Thomsen et al., 2010; Srinivasan et al., 2002; Stanulla et al., 2000; Thompson et al., 2005; Aaltonen et al., 2011). However, poor DNA yield and structural changes such as degradation and mutations may influence the results of both DNA sequencing and comprehensive marker analysis. Known factors contributing to possible poor DNA-quality are sample age, storage, and the fixation and staining methods (Falconi et al., 2007; Foss et al., 1994; Gilbert et al., 2007). In childhood ALL, bone marrow smears and biopsies are taken at diagnosis and several times during treatment, and these are stored for many years as part of the hospital's archival routines. Thus, there are often surplus samples representing a potential biobank and a possible means for obtaining DNA from the patients whose deaths are considered infectious TRDs. Unfortunately, there was no established method for multiple SNP sequencing based on DNA from such archival material, and the quality of such DNA was uncertain. We therefore wanted to test this in a separate study. Our research group has worked with a multiple SNP sequencing platform in other studies, and it was therefore natural to apply this SNP platform to the archival material. We also wanted to test the archival samples for other kinds of markers, and we selected an STR marker kit used routinely in our laboratory.

Poor DNA yield from archival samples may be an obstacle to multiple marker sequencing, and the techniques for DNA amplification are of interest. Whole-genome amplification (WGA) by multiple displacement amplification has been developed as a method for increasing the DNA content of a sample while keeping the sequence representation of the template intact (Spits et al., 2006). In a recently published study, our research group was successful in applying WGA to DNA from archival bone marrow samples (Lausten-Thomsen et al., 2010). Thus, we thought that WGA might represent a potential solution to the problem of poor DNA yield in our study.

Nordic Society of Paediatric Haematology and Oncology collaboration

NOPHO was established in 1981 in an effort to introduce uniformity in the treatment of children with leukaemia in the Nordic region, and common protocols for ALL treatment have been in use since 1992. NOPHO has gradually expanded its activities and areas of interest to include acute myeloid leukaemia, histiocytosis, pharmacology issues, idiopathic thrombocytopenic purpura, myelodysplastic syndrome and HSCT issues. Of the many working groups, the NOPHO Events Group focus on severe treatment-related events. As a participant in the NOPHO Events Group during the past five years, I have had the opportunity to participate in the ongoing surveillance and collection of data on TRDs in the last two ALL protocols (ALL-2000 and ALL-2008).

Personal motivation for the study

Clinical: As a paediatric oncologist, more than 10 years of daily observations of children with side effects from cancer treatment has been a major motivation to undertake this study.
Network: Being a member of a network with clinicians and researchers such as the NOPHO collaboration, with a special focus on toxic events through the NOPHO Events Group, has been very inspiring.
Registry: The ongoing voluntary data collection in NOPHO, originated by clinicians, provides the opportunity to approach daily clinical challenges on a large population-based scale with data collected in the NOPHO leukaemia registry in Stockholm.

What this thesis adds

This thesis adds new knowledge about the incidence, risk factors for and causes of TRDs in childhood ALL. The thesis also shows that host genomic SNP risk profiling of patients who develop infection during ALL treatment is possible and can identify important metabolic pathways related to susceptibility to infection during treatment. The challenges in using archival bone marrow samples as the DNA source for multiple biomarker analysis are also described. This work shows that clinical research is facilitated by international research networks, such as NOPHO, and that such international collaborations are feasible.

AIMS OF THE STUDY

- Because toxicity remains a great challenge in the treatment of ALL, one aim of the study was to explore the frequency of and risk factors for TRD in two consecutive NOPHO ALL protocols, ALL 1992 and ALL 2000. We also tried to compare differences in risk of TRD between the two protocols and to identify the causes of TRDs.

- Infection remains a significant problem in the treatment of childhood ALL, and only a few studies have reported on the genetic susceptibility to infection. Another aim was to explore the host genomic patterns that may influence susceptibility to infection during therapy. To achieve this aim, we expanded on our earlier study on the mannose-binding lectin gene (*MBL*) to include 34,000 target SNPs in candidate genes of possible relevance for childhood ALL treatment efficacy, immune-inflammatory response and toxicity. Using this approach and by exploring the relevant metabolic pathways, we hope to identify a genetic risk profile that can predict the risk of infection during induction treatment of childhood ALL.

- Studies on haematological malignancies to date, including those that used bone marrow samples as the DNA source have explored only a few biological markers (Kidas et al., 2009; Lausten-Thomsen et al., 2010; Pabst et al., 1996; Stanulla et al., 2000; Vince et al., 1998). To our knowledge, no studies have investigated thoroughly whether DNA from bone marrow samples is suitable for multiple STR and SNP analysis. Therefore, we aimed to evaluate whether archival bone marrow samples are suitable for multiple biomarker analysis.

MATERIAL AND METHODS

Study overview and design

The study was planned by Bendik Lund, Helge Klungland and Kjeld Schmiegelow, and was performed in close collaboration between the SNP/TOX Group (host genomics and toxicity) at Bonkolab,Copenhagen, the Center for Biological Sequence Analysis (CBS), Copenhagen, and the Genetic Research Group at the Department of Laboratory Medicine, Children's and Women's Health, Faculty of Medicine, Norwegian University of Science and Technology, Trondheim, Norway. In addition, Study I was performed in close collaboration with the NOPHO Events Group.

The three studies used different designs. Study I was *retrospective* in the sense that detailed clinical information about the causes of death (TRDs) was collected years after the deaths occurred. By contrast, some of the clinical information (anthropometric data, clinical and laboratory data, treatment response, events) was collected *prospectively* during monitoring of the respective ALL protocols. Study II was a clinical retrospective case-control gene-association study of infectious events. The clinical information used in this study was collected retrospectively during an earlier study on *MBL*-gene polymorphism by our research group (Lausen et al., 2006). SNP analysis was performed in 2010 and 2012 after the patients had completed their treatment. Study III was an experimental laboratory study of archival bone marrow samples that compared biomarkers in bone marrow samples with those in the corresponding blood samples.

Patients and samples

Patients

The participants in the present studies were childhood ALL patients aged 1.0 to 17.9 years at diagnosis who were treated according to the common Nordic ALL protocols NOPHO ALL-1992, ALL-2000 and ALL-2008. The ALL-1992 protocol was in use from 1992 to 2000, the ALL-2000 protocol from 2000 to 2008, and the ALL-2008 protocol started in 2008 and is still running. About 90% of all children with ALL diagnosed within the Nordic region (Iceland, Finland, Sweden, Denmark and Norway) are included in the running protocol and have agreed to allow the data to be collected and registered in the NOPHO ALL registry at Karolinska Institute, Stockholm. As of 31/12/2012, 3,488 childhood ALL patients have been included in the registry (all three protocols) ("NOPHO Annual Report 2013," 2013). The

patients were stratified into either a low-risk group (standard or intermediate intensity) or high-risk group and were treated according to the protocol that was current at the time. This involved chemotherapy treatment for 2-2.5 years in all protocols (Schmiegelow et al., 2010; Toft et al., 2013). The treatment details are outlined in Papers I and II.

Papers I and II were based on data from the NOPHO ALL registry. Additional clinical data on the causes of death (TRD) were collected through a questionnaire prepared as part of the work of the NOPHO Events Group and was collected on behalf of the NOPHO ALL Registry (Appendix). As part of another study by our research group (*MBL*-study) (Lausen et al., 2006) additional clinical data were collected retrospectively from the patient files.

Paper I included patients from the ALL registry between 1992 and 2008 (ALL-1992 and ALL-2000 protocols) in which 2,735 childhood ALL patients constituted the study cohort. Of these, 88 TRDs were identified and served as cases in the study.

Paper II included patients treated according to the NOPHO ALL-1992 protocol at Rigshospitalet, Copenhagen, 1992 to 2000. One hundred and thirty-seven patients fulfilled these criteria and were included in the previously mentioned study on *MBL* polymorphisms in relation to infection during the first 50 days of antileukaemic therapy (Lausen et al., 2006). Blood samples for extended SNP profiling were available for 69 patients, who constituted the actual study cohort.

Paper III included samples from 17 randomly selected Norwegian and Danish patients included in the NOPHO ALL-2000 and ALL-2008 protocols.

Grouping of patients

In Study II (SNP associations) patients were grouped twice. The first grouping included patients who had at least one infectious event (n=48) during the seven-week induction period, and the remaining 21 infection-free patients served as controls. The second grouping included patients from the first grouping who also had at least one positive blood culture (n=23), and the remaining 46 patients served as controls. Culture-positive patients were regarded as having a potentially more severe infectious episode.

Blood samples and archival bone marrow samples

All blood samples used in the study were obtained during morphological remission and were sampled between 1992 and 2010. The archival bone marrow samples were collected from the archives at Rigshospitalet, Copenhagen, and St. Olavs Hospital, Trondheim. The samples

included both Giemsa-stained bone marrow smears and formalin-fixed paraffin-embedded (FFPE) bone marrow biopsies. They had been stored at room temperature for 0-10 years. Only surplus material was used.

Candidate genes and SNP selection

In our gene-association study (Paper II) and evaluation of multiple markers from archival samples (Paper III), we chose an extended candidate gene approach. Through an expert panel co-ordinated by the SNP/TOX Group at Bonkolab, non-systematic selection of about 2,350 candidate genes with possible relevance for childhood ALL was achieved. The major focus for gene selection was treatment efficacy and toxicity, and the following areas/domains were included: i) pharmacogenetics, ii) immunogenetics, iii) apoptosis, iv) organ-specific toxicities (including thrombosis/bleeding, vincristine neuropathy and pancreatitis), v) cell cycle control genes and vi) DNA repair and mitosis. The targeted SNPs within these candidate genes were selected based on the influence on their transcript (according to Ensembl (Hubbard et al., 2002) annotations) as: i) non-synonymous coding, ii) frame-shift coding, iii) regulatory region, iv) stop loss, v) stop gain, vi) splice site, vii) within non-coding genes, and viii) within mature micro-RNA. This resulted in selection of about 34,000 targeted SNPs. Our research group had earlier developed a cost-effective next-generation sequencing capture assay for SNP analysis that allows pooling before capture and simultaneous genotyping of several samples for a large number of variants (Wesolowska et al., 2011), and this assay was used for our multiple SNP analysis.

Whole-genome amplification

WGA is a method that amplifies small amounts of the DNA template giving a higher DNA yield with conserved sequence representation. A new method of WGA called *multiple displacement amplification* has been developed and generates DNA with higher molecular weight and better genome coverage (Spits et al., 2006). The method makes use of a DNA polymerase (Φ29) and is based on annealing of random hexamers followed by strand displacement, as shown in Fig. 5.

The same method had been used earlier for amplification of DNA from bone marrow smears in our research group (Lausten-Thomsen et al., 2010). To overcome a possible

problem with poor DNA yield from our archival samples, WGA was introduced into our laboratory work-flow after DNA extraction and before SNP/STR analysis.

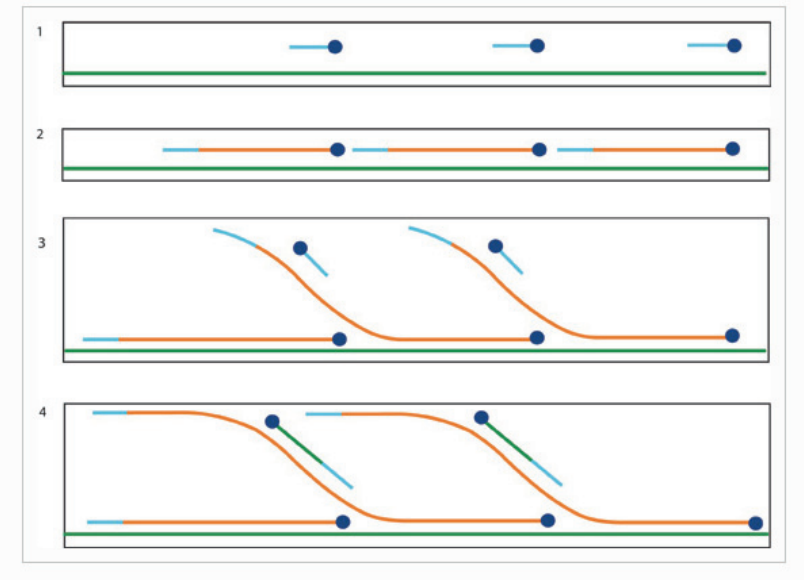


Figure 5. Overview of the principle of multiple displacement amplification (from paper: *Whole-genome multiple displacement amplification from single cells* (Spits et al., 2006)).

1) The random hexamers (represented by a blue line) bind to the denatured DNA (green line).
2) The Φ29 DNA polymerase (blue circle) extends the primers until it reaches newly synthesized double-stranded DNA (orange line).
3) The enzyme proceeds to displace the strand and continues the polymerization, while primers bind to the newly synthesized DNA.
4) Polymerization starts on the new strands, forming a hyper-branched structure.

Laboratory work

The laboratory work was performed in Trondheim (LBK/NTNU) and Copenhagen (Bonkolab and CBS). Samples were sequenced at commercial sequencing facilities at BGI in Hong Kong and at Aros in Århus.

Preparation of samples

Blood samples

DNA from blood was extracted and purified either by standard sodium chloride and ethanol precipitation or by using the QIAamp® DNA Mini Kit (Qiagen, GmbH, Hilden, Germany) according to the manufacturer's instructions.

Archival bone marrow samples

The bone marrow smears were observed macroscopically and microscopically to visually ensure the required minimum cell content. Among the 21 included smears, 12 had cover slides, which were separated from their glass slide by immersion in xylene for 2-4 days, after which the samples were washed three times with 95% ethanol and air-dried for 2-4 days. PBS buffer (0.5 ml) was pipetted onto the glass slide and the cells were carefully scraped off the slide surface into a microcentrifuge tube. DNA was extracted using a QIAamp® DNA Micro kit (Qiagen) according to the manufacturer's instructions and stored at – 20 °C. From these 21 DNA samples, 51 aliquots were taken for further analysis. The REPLI-g® FFPE kit (Qiagen) was chosen for WGA because it is designed specifically for fragmented DNA. Amplification was performed according to the manufacturer's instructions. For the bone marrow biopsies, a standard microtome technique was used to prepare new sections of FFPE tissue with a thickness of 10 μm; five sections were prepared per sample. DNA was processed by two methods: i) for marker analysis without WGA, DNA was extracted using the QIAamp® FFPE tissue Kit (Qiagen) according to the manufacturer's instructions, and ii) for marker analysis including WGA, the REPLI-g® FFPE kit (Qiagen) was used; this kit includes DNA extraction, ligation and amplification in the same procedure.

Quantity and quality measurements

Because of the expected poor quality and yield of DNA extracted from the archival material, quality and quantity assessment was considered a crucial step. To measure DNA concentration, both spectrophotometry (SP) (NanoDrop® ND 1000 Spectrophotometer,

NanoDrop Technologies, Inc., Wilmington, DE, USA) and fluorometry (FM) (Qubit®, dsDNA BR assay, Invitrogen, Carlsbad, CA, USA) were used. Gel-electrophoresis was used to estimate fragment lengths of both amplified and non-amplified material.

STR marker analysis

For a subset (n=11) of the patients, STR marker analyses were performed on a total of 31 archival samples. The AmpFℓ STR® Profiler® kit (Applied Biosystems, Foster City, CA, USA) was used. Amplification was performed according to the manufacturer's recommendation. The 10 STR loci had lengths varying from 107-113 bp (amelogenin, gender marker) to 281-317 bp (CSF1PO marker). Marker profiles of the archival material were compared with the corresponding marker profiles from blood samples. The detection rate of STRs was defined as the total number of detected STRs from the particular archival samples for all included patients in that group divided by the total number of STRs detected in blood samples for the group. This measure takes into account that some patients were either homozygous or heterozygous for the respective marker.

Library preparation

The library was prepared according to the SureSelect Target Enrichment System protocol April 2009/November 2010 (Agilent Technologies, Santa Clara, CA, USA) as described earlier (Wesolowska et al., 2011). Briefly, 3 µg of genomic DNA was sheared in a Covaris S2 System (Covaris, Inc., Woburn, MA, USA), and the DNA was purified and subjected to end-repair. For multiplexing, patient-specific barcodes of four bases were ligated to the DNA fragments. The library samples were then pooled in groups of 3-8 samples and hybridized to custom-designed baits (Sure Select Oligo Capture Library, Agilent Technologies), with two baits targeting each SNP (Fig.6). After PCR amplification, the library was sequenced using an Illumina HiSeq 2000 system (Illumina, Inc., San Diego, CA, USA).

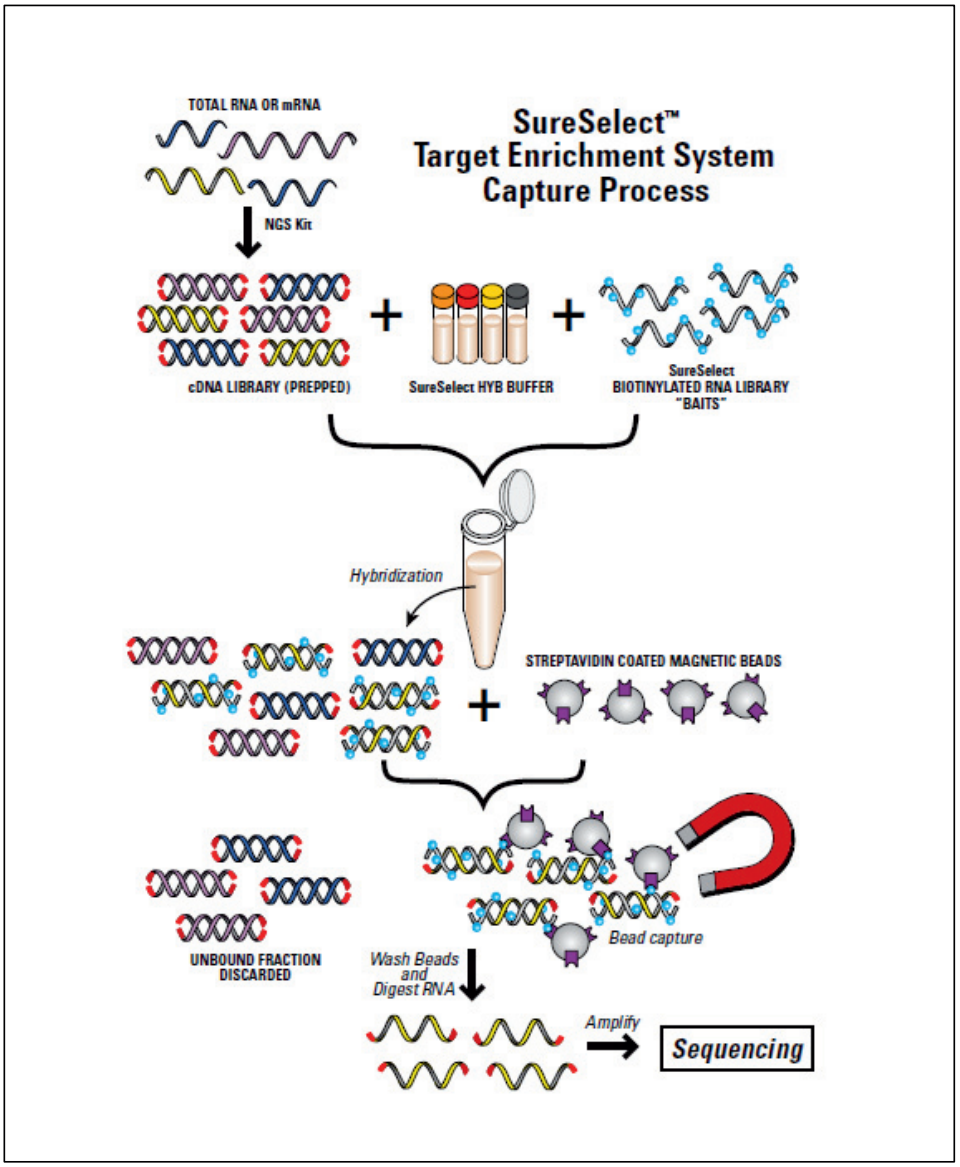


Figure 6. Work flow of Agilent's SureSelect Target Enrichment System. The patient's DNA is fragmented by ultrasound into shorter fragments (Covaris) and hybridized with the biotinylated library of RNA capture baits. Hybridized RNA-DNA duplexes are pulled down with streptavidin-coated magnetic beads. After washing the beads, the RNA baits are digested, and the subset of initial DNA is ready for sequencing. Source: www.genomic.agilent.com.

Outcome measures

Paper I: Number and percentages of TRDs, causes of TRD, clinical parameters and risk factors including WBC, high- and low-risk patients, Down syndrome, CNS-disease and HSCT.

Paper II: Number of patients with infectious events and a positive culture; significant SNPs and pathways associated with infections; classification and regression tree (CART) analysis for predicting infections including SNP profiling and pathway analysis; time-to-an-infectious-event.

Paper III: DNA concentration measurements; STR detection rate; SNP call rates and concordance.

Ethics

The study was approved by research ethics committees in Denmark (H-D-2007-0100) and Norway (REC-nr 2010/803). Written informed consent was obtained from all patients and/or their parents. For the treatment protocols and registration of data in the NOPHO ALL registry in Stockholm, informed consent had been collected for all participants. The treatment protocols were approved by the national or regional ethics committees in the five Nordic countries and the respective national childhood leukaemia groups. The study was conducted in accordance with the Declaration of Helsinki.

Statistics and bioinformatics

Common statistics

All statistical analyses excluding the bioinformatics were performed by Bendik Lund. In Paper I, Stian Lydersen supervised the statistical analysis. The software package SPSS Statistics (versions 16-20; SPSS, Inc. Chicago, IL, USA and IBM Corp., Armonk, NY, USA) was used. Proportions were compared using the chi-square test. Kaplan-Meier plots and survival tables were used for time-to-event analysis, and subgroups were compared using the log-rank test. In Study I, TRD was used as the main event. Other events (resistant disease, relapse or SMN) were censored at the time of these events. Time to TRD was defined as the

time from the diagnosis of ALL to the time of death. Patients who experienced no events were censored at the time of the last follow-up. Cox proportional-hazard regression analysis was performed for multivariate time-to-event analysis. In Paper I, the following covariates were included: sex, age (< or ≥10 years), WBC (< or ≥200 $\times 10^9$/L), B-cell precursor vs. T-cell disease, protocol (ALL-1992 vs. ALL-2000), presence or absence of CNS-disease, Down syndrome and HSCT in first complete remission (CR1). For patients who underwent HSCT during CR1, a time-dependent covariate was defined as "0" before, and "1" after the date of HSCT; for all other patients, this time-dependent covariate was defined as "0". In Paper II, covariates included ANC and age.

Bioinformatics

All bioinformatic analysis was performed by Agata Wesolowska-Andersen (Papers II and III) at CBS/DTU. Briefly, the sequencing reads for each sample were mapped to the reference human genome build 37 (GRCh37) using Burrow-Wheelers Alignment algorithms (Li & Durbin, 2009; Li et al., 2010). The alignment was refined by means of the alignment score, base quality score recalibration and around indel realignment using Genome Analysis Toolkit (McKenna et al., 2010), and duplicate sequences were removed with Picard Tools' (Picard) MarkDuplicates function. SNP calling was performed with the SAMtools package (Li, Handsaker, et al., 2009) using the default settings. The variants were filtered using the vcfutils.pl script from the SAMtools package to remove any bias attributable to base quality score, strand or position along the read. The high quality of SNP calling with our multiplexing technique has been validated with various PCR methods (Wesolowska et al., 2011). Variant annotation was performed with Ensembl Variant Effect Predictor script (McLaren et al., 2010). The data were converted into .ped and .map file formats readable by PLINK (Purcell et al., 2007) using vcftools (Danecek et al., 2011).

Bioinformatics specifically related to Paper II: The associations between single SNPs and the risk of infection and of having a positive culture were analysed by Fisher's exact test implemented in PLINK. Only SNPs with an observed MAF >5% and ≥50% of non-missing genotypes >10× sequencing depth were included in the analysis. A non-strict cut-off of 50% of non-missing genotypes was chosen because of the novelty of this genotyping-technique, which is often complicated by the presence of genomic regions that are difficult to target with capture baits or are difficult to sequence. The obtained *p*-values were adjusted for multiple testing with up to one million adaptive permutations (Besag et al., 1991). Finally, the adjusted

p-values were plotted on quantile-quantile plots using publicly available R script (http://gettinggeneticsdone.blogspot.com/) and on a Manhattan plot using CIRCOS software version 0.52 (Krzywinski et al., 2009). All potentially functional (non-synonymous coding, frame-shift coding, stop codon and splice site) SNPs genotyped in this study with MAF > 0.01 residing in the pathways genes were retrieved for Reactome pathways (Joshi-Tope et al., 2005) excluding the top two pathway levels. The number of SNPs per pathway ranged from 1 to 82, and each SNP was encoded by three values between 0 and 1 corresponding to the likelihood of each genotype calculated from the VCF file produced by SAMtools (Li, Handsaker, et al., 2009). Missing genotype calls were encoded as observed population frequencies for the three genotypes. Associations with an infectious event and positive blood culture were identified by training artificial neural networks on subsets of SNPs from each pathway with threefold cross validation. For each pathway, all combinations of up to three SNPs were assessed, and the combinations of SNPs tested were ranked by Matthew's correlation coefficient (MCC) calculated as:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

where TP is the number of true positives, TN the number of true negatives, FP the number of false positives and FN the number of false negatives. For each resulting best combination of SNPs for each pathway, the neural network parameters were optimized by testing multiple settings of hidden neurons and training cycles. The pathways were then ranked by the MCC of the best combination of SNPs for each pathway. For all best combinations, the area under the receiver-operating characteristic curve (AUC) was calculated with 95% confidence intervals with 'pROC' R package. CART analyses were performed using rpart R package applying threefold cross-validation (Therneau et al., 1997) using the genotypes of the SNPs associated with the risk of infection and positive culture with *p*-values <0.01 together with the patients' age, sex, risk group, immunophenotype and WBC at diagnosis. The same analysis was repeated by also including pathway-based predictions achieved from top 10 artificial neural network classifiers from the pathway analysis.

Bioinformatics specifically related to Paper III: For the final analysis, SNP call rates were calculated, and archival SNP profiles were compared with corresponding blood SNP profiles for concordance estimations.

RESULTS

Summary of papers

Paper I

Title

Risk factors for treatment related mortality in childhood acute lymphoblastic leukaemia

Aims of Paper I

The aim of this paper was to investigate the pattern of TRDs and possible risk factors in the NOPHO ALL-1992 and ALL-2000 protocols. Fifty-five TRDs were identified among the 1,645 ALL-1992 patients and 33 among the 1,090 ALL-2000 patients.

Main results

Of the total of 2,735 patients, 51 girls and 37 boys (3.2%) with a median age of 4 years (75% range 2.0-11.0 years) at diagnosis experienced a TRD. The TRDs comprised 25% of all 354 deaths in the study population, with 240 deaths after relapse as the largest group. The incidence of TRDs did not differ between the two protocols (3.4% vs. 3.2%, ALL-1992 and ALL-2000, respectively). For the risk groups (stratification based on upfront criteria), the cumulative incidence rates of TRD were 1.7 ± 0.4%, 2.4 ± 0.5% and 6.7 ± 0.9% for the standard-risk (SR), intermediate-risk (IR) and high-risk (HR) groups, respectively ($p < 0.001$). Five patients died before initiation of anti-leukaemic therapy (0.2%), and the overall subsequent risks of induction death and death in CR1 were 1.2% and 1.8%, respectively. Infections were the major cause of death and accounted for 72% of all cases, including nine deaths from *Pseudomonas aeruginosa* and 11 deaths from fungal infections. Other causes of death included bleeding or thrombosis (eight patients), tumour burden-related toxicity (seven patients) and organ toxicity (seven patients). The independent risk factors for TRD were identified as female gender (hazard ratio (HR): 2.2, 95% confidence interval (95% CI): 1.4-3.4), high WBC ($\geq 200 \times 10^9$/L) at diagnosis (HR: 3.5, 95% CI: 1.7-7.1), T-cell disease (HR: 1.9, 95% CI: 1.01-3.7), Down syndrome (HR: 7.3, 95% CI: 3.6-14.9), and HSCT in CR1 (HR: 8.0, 95% CI: 3.3-19.5).

Paper II

Title

Host genome variations and risk of infections during induction treatment of childhood acute lymphoblastic leukaemia

Aims of Paper II

The aim of this paper was to investigate the associations between host genomic variations and risk of infections during induction treatment of childhood ALL.

Main results

Forty-eight (70%) patients experienced at least one infectious event during induction treatment, of which 23 patients had at least one positive blood culture. Infectious events and a positive blood culture were significantly associated with 24 and 21 SNPs, respectively ($p <$ 0.01). CART analysis identified rs11033797 (*OR51F1*), rs2835265 (*CBR1*), rs28627172 (*POLDIP3*) and rs1129844 (*CCL11*) as predictive of outcome. Among the 61 patients whose readouts were available for all four SNPs, 40 of 41 SNP profile-positive patients experienced at least one infectious event compared with five of the remaining 20 patients (HR 9.0, 95% CI: 3.4-23.5, which was unchanged after adjustment for neutrophil count). Pathway analysis identified variations in "G-protein-coupled receptor (GPCR) downstream signalling", "Bile acid and bile salt metabolism" and "Class I MHC mediated antigen processing and presentation" to be highly predictive of infections.

Paper III

Title

Archival bone marrow samples: suitable for multiple biomarker analysis?

Aims of Paper III

The aim of this paper was to evaluate the usability of archival bone marrow smears and biopsies for DNA extraction and purification, WGA, multiple marker analysis including 10 STRs, and the comprehensive genotyping of 33,683 SNPs with multiplexed targeted next-generation sequencing. A total of 73 samples from 21 bone marrow smears and 13 bone marrow biopsies from 18 Danish and Norwegian patients with childhood ALL were included and compared with the corresponding blood samples.

Main results

We found that the measurement of DNA concentration after DNA extraction was dependent on the detection method and that SP overestimated DNA amount compared with FM. The mean DNA concentrations were 68.0 ± 43.5 ng/µL and 29.0 ± 20.0 ng/µL for SP and FM, respectively ($p < 0.001$) for smears, and 103.8 ± 73.1 ng/µL and 5.1 ± 3.7 ng/µL, respectively, ($p = 0.003$) for biopsies. For the STR-marker analysis, all 10 markers were identified in all blood samples. The detection rate of markers for smears was 93% compared with 88% for biopsies. Analysis of WGA-products in smears and biopsies showed significantly lower detection rates of 28% ($p < 0.001$) and 63% ($p < 0.001$), respectively. When samples were grouped according to time after collection, the detection rate of markers was significantly higher for samples stored for 0-3 years than for samples stored for 4-10 years. In the STR analysis, the detection rate declined slightly for longer fragments. Following WGA, this decrease was more pronounced. Of the initial 42 archival samples taken for multiple SNP analysis, 13 samples (all derived from smears) had SNP calls >10,000 at 4× read depth. Of these, seven archival samples had SNP calls >17,000 (i.e., 50% of about 34,000 intended SNPs; Table III, Paper III). Concordance varied from 87% to 100% for total reference SNP sites and, except for two samples (ID1 and ID5), 73% to 99% for total variant SNP sites. The seven samples belonged to group 1 (n = 2; 0-3 years old, +WGA), group 2 (n = 2; 0-3 years old, -WGA) and group 3 (n = 3; 4-10 years old, +WGA). The two smears with highest concordance (96-100%) belonged to groups 1 and 2. None of the eight samples derived from biopsies had SNP calls >10,000.

DISCUSSION

Main findings

Among the 2,735 patients included in Study I, the overall TRD rate was 3.2%, indicating that treatment-related severe toxicity still represents a major challenge in childhood ALL. Because the number of patients dying from leukaemia is decreasing, TRDs comprise an increasing proportion of overall mortality. We identified tumour burden, infection, bleeding/thrombosis and specific organ toxicity as causes of TRD, with infection as the most common cause. To reduce toxicity and prevent TRD, more focus is needed to investigate the overall risk factors for outcome, i.e., the tumour burden, the treatment intensity, and the supportive care. There is also the need for increased knowledge about the relevant host genetic factors. Multiple regression analysis identified female gender, high WBC at diagnosis ($\geq 200 \times 10^9$/L), T-cell disease, Down syndrome and HSCT in CR1 as risk factors for TRD.

In our study on host genetic risk factors for infectious events during induction treatment, we identified a SNP risk profile that could predict the risk of an infectious event with 98% accuracy among SNP profile-positive patients. In addition, metabolic pathways related to the risk of infectious events were identified suggesting the possible underlying mechanisms involved in the development of infections among ALL patients. We also demonstrated a potential model for further research in this area and found that genetic risk profiling may be possible even in studies with a relatively small sample size.

When investigating the use of archival material for multiple marker studies, we found that archival bone marrow smears and biopsies can be used for DNA extraction and purification. The importance of the method used to measure DNA concentration measure method was illustrated by our findings that SP overestimated DNA concentration compared with FM. Multiple STR analysis was successful for both smears and biopsies, but older samples had a lower success rate compared with younger samples. We also found that multiple SNP array analysis from archival samples may be possible, but this method was successful for only a few samples. Hence, the method must be optimized.

Main findings in relation to other research

TRD: treatment intensity vs. toxicity

Our finding of an overall TRD rate of 3.2% is similar to that found in other studies (Blanco et al., 2012; Prucker et al., 2009). The TRD rate did not change significantly between the ALL-1992 protocol and the ALL-2000 protocol. This fits well with a review by Blanco and co-workers showing that the TRD rate has not changed much over the past 30 years (Blanco et al., 2012).

Successful treatment of childhood cancer implies a fine balance between treatment intensity and side effects. In our study, TRDs accounted for 25% of the 354 deaths in the study population. Most of the other deaths occurred after relapse. The percentage of deaths caused by relapse or treatment-related toxicity might be age-dependent as shown by Pichler et al. (Pichler et al., 2013). They found that the inferior overall prognosis in adolescents was caused by a higher TRD-rate and not by relapse, indicating either that adolescents more often receive intensive treatment or that adolescents are less able to tolerate treatment compared with younger children. We found that age (10.0-14.9 vs. 1.0-9.9 years) was not a risk factor for TRD, but we did not compare between deaths after relapse and TRD.

Having a relapse indicates more chemotherapy-resistant disease and significantly decreases the chance of a cure (Bhojwani et al., 2013; Nguyen et al., 2008). High treatment intensity reduces the chance of a relapse, but increases the risk of severe side effects, and vice versa, as illustrated in Fig. 7.

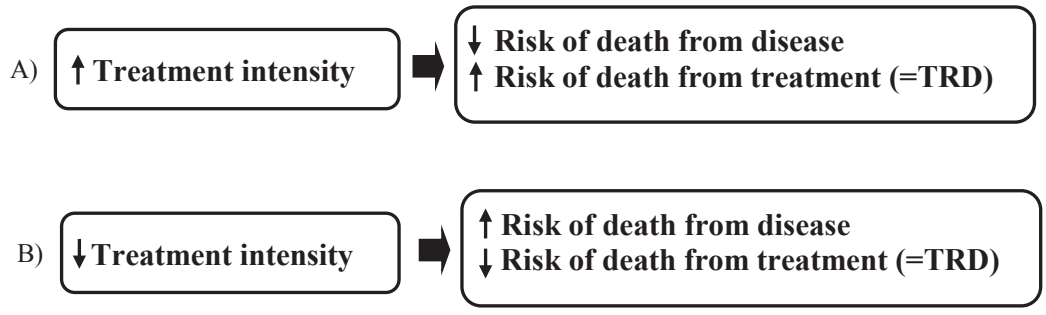


Figure 7.

This concept is consistent with our observation (Paper I) that the cumulative incidence of TRD was higher in those who received more intensive treatment: 1.7%, 2.4% and 6.7% for the SR, IR and HR groups, respectively.

A great challenge for clinicians is to use treatment at an intensity that is required for disease control and to avoid relapse while at the same time not overtreating patients and causing intolerable side effects. The question is: how to allocate the individual patients to the most appropriate treatment? At diagnosis and after response evaluation, the following is only partly known: i) which patients will respond to the chemotherapy prescribed in terms of disease control (blast cell chemotherapy sensitivity, relapse risk), and ii) which patients have host genetic factors that make them more vulnerable to side effects. As outlined in Paper I, risk-adapted therapy is based on different parameters related to the disease, such as the WBC at diagnosis, immunophenotype (T- or B-cell), CNS/testicular involvement, cytogenetic alterations, and the response to the initial therapy. However, some patients experience TRD even when having low-intensity treatment (SR and IR groups). Hence, for these patients, the prescribed low-intensity drug-combinations are in fact too intensive. An example of incorporating the host pharmacogenetics into drug dosing is the use of *TPMT* SNP-genotyping for 6-MP dosing. This was not part of the ALL-1992 protocol, but was incorporated in the ALL-2000 protocol and has been continued in our ongoing ALL-2008 protocol. This approach is also consistent with modern pharmacological thinking about personalised medicine. The paradigm of "all patients same drug and dose" seems not appropriate in terms of the increasing knowledge about host genetics and disease biology (Blay et al., 2012; Davidsen et al., 2008; Hudson, 2013) .

TRDs occurred at different phases of therapy. Seven patients died from tumour burden-related problems within the first days after the diagnosis, including five patients who died before the start of any anti-leukaemic treatment. Handling of such patients is challenging and the outcome is clearly related to the rapidity of the diagnostic work-up and initiation of therapy. Modern treatment with recombinant urate oxidase (rasburicase) has been shown to be effective for prophylaxis and treatment of urate-related problems of tumour lysis syndrome in childhood malignancies (Pui et al., 2001). This makes it possible to initiate anti-leukaemic treatment rapidly with full dose GC therapy after administration of rasburicase. This approach might have prevented the tumour burden-related TRDs seen in this study. One can argue whether these patients should be classified as TRDs. If "treatment" means "chemotherapy" or "GC-treatment" these cases might instead be classified as "disease-related death". However,

these patients are challenging for the treating physician and highlight the clinical spectrum of severely ill ALL patients. Exploring such patients and their causes of death focuses attention to better handling the first days after the diagnosis.

Treatment-induced immunosuppression

Treatment-induced immunosuppression includes neutropenia, impaired humoral antibody response, impaired cell-mediated immunity, phagocytic defects and disturbed cytokine functions (Eyrich et al., 2009; Hann et al., 1997; Pui, 2006). All of these factors contribute to the increased risk of infection. Our findings that 70% of patients experienced at least one infectious event during induction treatment and that the first event happened relatively early after the start of treatment (median time to infection, 2.5 days; Paper II) are consistent with the relationship between these factors and the risk of infection. Chemotherapy-related risk of infections has been well documented by others (Afzal et al., 2009; Graubner et al., 2008; Stabell et al., 2008). In our study, infections were the most common cause of TRD (PaperI), a finding that has also been reported by others (Conter et al., 2000; Hargrave et al., 2001; Prucker et al., 2009; Wheeler et al., 1996). Hence, there is a widespread use of antibiotics both in infectious events (with or without neutropenia) and as prophylaxis. In cases of febrile neutropenia or any suspected infectious event, empirical therapy is based on the most probable causative organisms and their likely susceptibility patterns. Different antibiotic regimens are effective, many of which include either a combination of penicillin and an aminoglycoside or, as monotherapy, a broad-spectrum β-lactam such as a carbapenem (Stabell et al., 2008; Yildirim et al., 2008).

Prophylactic antibiotics are used to varying degrees against bacteria. A review by van de Wetering et al. (van de Wetering et al., 2005) showed that oral prophylactic antibiotics decreased G-negative bacteraemia and infection-related mortality in oncology patients during neutropenic episodes. However, only three of the 22 trials reviewed included paediatric patients. Larger prospective studies are needed to explore whether antibacterial prophylaxis during the neutropenic phases in the first three months of anti-leukaemic therapy can reduce the risk of infectious TRDs in patients with childhood ALL.

A well-known and potentially fatal infection in immunocompromised patients is *Pneumocystis jirovecii* pneumonia (PCP). Trimethoprim-sulfamethoxazole (TMP/SMX) has been shown to be an effective prophylactic against PCP (Green et al., 2007), but there is no international consensus regarding the dosing of TMP/SMX prophylaxis used in childhood malignancies. Giving the drug three days a week seems as effective as giving it daily, as

shown in a study by our research group (Levinsen et al., 2012). In our present study (Paper I), only one patient died from PCP. There was no uniform approach to prophylaxis against PCP infection during the study period, and no patient-specific data on prophylaxis against PCP were available. Thus, it is unknown to what extent local strategies for PCP-prophylaxis may have contributed to the low frequency of mortality from PCP in our study.

The finding that 76% of the TRDs occurred in the first 80 days after the start of treatment, including nine of the 11 deaths due to fungal infection, shows that this treatment period is the most crucial with regard to the risk of severe infection and immunosuppression, mainly caused by the initial tumour load and very intensive chemotherapy treatment including GCs. The fact that 52 % of the patients who experienced TRD because of infection were not neutropenic during the week preceding death shows that the risk of severe infection not only is associated with neutropenia, but also with suppression of other parts of the immune system.

Some infectious deaths occurred despite adequate antibiotic treatment according to the resistance pattern of the microorganism(s) in question. This raises the question of a possible role of critical host genetic factors including pharmacogenetics (Davidsen et al., 2008; Karathanasis et al., 2009) and immunogenetics. Studies on genetic polymorphisms in immunoregulating mediators have shown an association with outcome during childhood leukaemia induction therapy (Kidas et al., 2009), in childhood AML patients (Lehrnbecher et al., 2005), in childhood malignancies in general (Neth et al., 2001) and in post-HSCT patients (Middleton et al., 2002; Shamim et al., 2006). However, most studies are candidate gene-based involving only a few genes, and genome-wide studies of variations within the immune response involving multiple genes and haplotypes are lacking. Identification of possible immunogenetic risk profiles at the start of therapy might be helpful in the development of risk-adapted and individualized supportive care.

Of the 88 TRDs, 15 (17%) were caused by specific organ toxicities or haemostasis-related causes. Most organ-specific or haemostasis-related toxicities, such as pancreatitis, liver-toxicities and thrombosis or bleeding, are treated successfully by careful supportive care; therefore, general conclusions about these small subgroups cannot be drawn. Infections seem to be the most critical issue relating to TRD and exploring host genomics seems reasonable, as others have indicated (Graubner et al., 2008).

SNP risk profiling

To some extent, it remains unknown why some patients develop severe infections during treatment while others, even those within the same sex, age and risk group, experience only mild infections. We were therefore interested in exploring the possible role of inherited genetic factors. First, we focused on identifying a SNP profile predictive of infectious events regardless of its biological relevance. Second, we focused on genes and pathways of relevance for infections in children with ALL to increase our understanding of the underlying biological mechanisms.

Our findings that 24 and 21 SNPs were associated with infectious events and positive blood cultures, respectively, (Fig. 8, from Paper II), strongly suggest that common host genomic variants within immune function, drug disposition and other intracellular signalling mechanisms play a critical role in the risk of susceptibility to infection in children with ALL.

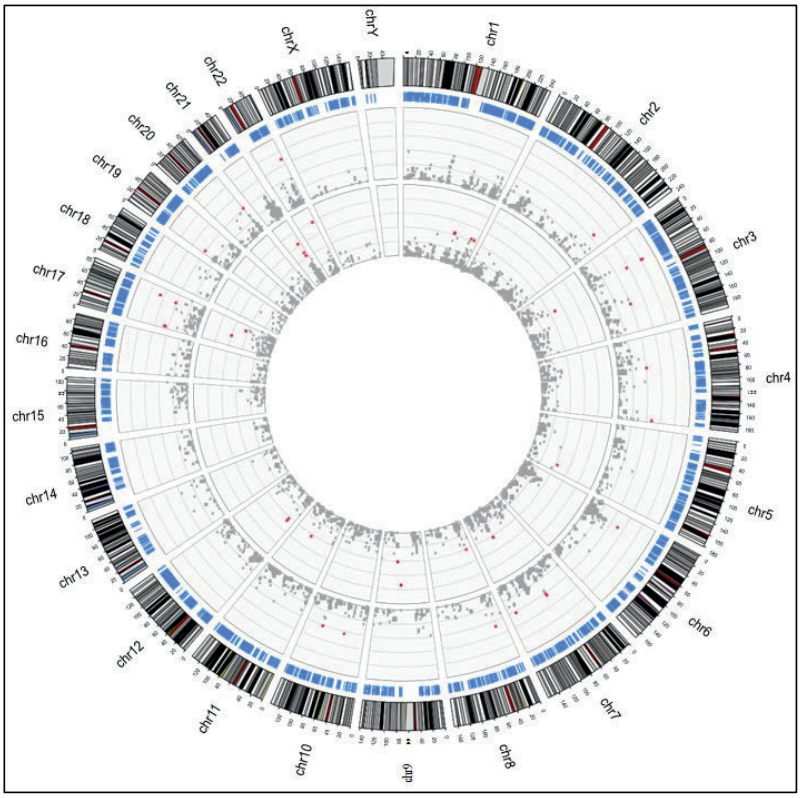


Figure 8. (Paper II)**.** Manhattan plot showing the association of SNPs with infectious events (inner ring of blocks) and positive culture (outer ring of blocks). The radius represents the $-\log_{10}$ of *p*-values, and each SNP is plotted based on its position in the chromosome and as a function of its *p*-value. Red dots represent SNPs with *p*-values <0.01. Grey dots represent SNPs with *p*-values >0.01. Blue tick marks represent the positions of baits used for target capture.

Using CART analysis, we identified a four-SNP risk profile based on our primary findings of 24 SNPs associated with the presence of an infectious event. For patients termed "SNP profile positive", 40 of 41 patients experienced an infectious event compared with only five of the 20 "SNP profile-negative" patients. To our knowledge, multiple SNP-profiling has not been used to predict the risk of infection in childhood ALL patients. SNP data and CART analysis has been used in only a few studies of childhood ALL, but with other end-points such as ALL susceptibility and risk of osteonecrosis (Chan et al., 2011; Kawedia et al., 2011). CART analysis is a mathematical tool helpful for identifying the most predictive SNPs related to outcome. The SNPs included in the profile do not necessarily belong to the same pathways,

but are mathematically selected as the most predictive ones. The significant SNPs might also not be causative in pathways related to infections, but may serve as markers (tag-SNP) because of linkage disequilibrium with other causative SNPs. The simplicity of CART-analysis makes the risk profile data more comprehensible and clinically applicable than complex pathway interactions.

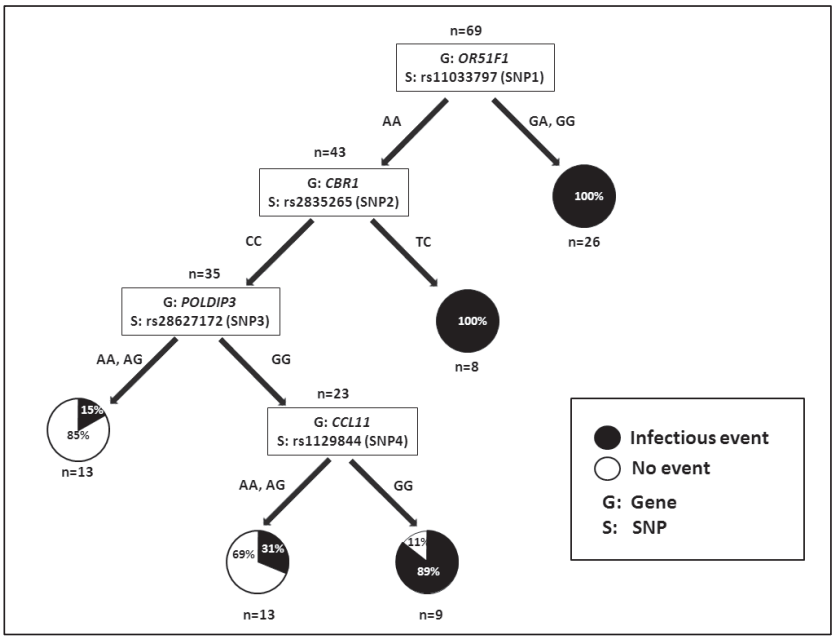


Figure 9. CART (classification and regression tree) diagram for "infectious event" showing the most predictive SNP-sequences and the corresponding genes (Paper II).

The four predictive SNPs in our CART analysis (Fig. 9, from Paper II) link four genes to the risk of having an infectious event and two genes to the risk of a positive culture. The SNPs and genes, and their relationships with ALL, are summarized in Table II. To our knowledge, the olfactory receptor *OR51F1*-gene has not been related to infections, childhood leukaemia or chemotherapy side effects. It was included in our multiple SNP analysis platform because it resided in an intron of another gene, *MMP26*, which was selected because

of its relationship with tumour invasion and metastasis. Some theoretically functional analogy between germinal centre B-lymphocytes and olfactory sensory neurons might exist (Otaki et al., 2005), but one can only speculate whether there is a hitherto undetected link between *OR51F1* and the risk of infection. It might also be a false-positive finding reflecting mutational heterogeneity frequently seen, particularly for olfactory receptors (Lawrence et al., 2013). The *CBR1*-gene is associated with metabolism of anthracycline (Bains et al., 2009), which is part of the induction treatment. Patients with reduced metabolism of doxorubicin might experience increased toxicity, causing prolonged neutropenia and thereby an increased risk of infection. An association with infection in immunocompromised patients has not been reported for the *POLDIP3*-gene, but it has been associated with sleep patterns in childhood ALL patients during dexamethasone treatment (Vallance et al., 2010). The *CCL11*-gene relates to asthma, allergic diseases and parasite infections (Simons et al., 2005), but as an inflammatory signalling molecule, it might play a role in host defence against infection in childhood ALL patients.

For the other end-point in our study, culture-positive patients, the most predictive SNP was located in the *FLNB*-gene, which encodes filamin B, a cytoplasmic protein that binds actin and connects plasma membrane to the intracellular cytoskeleton, thereby changing the cell shape. It is expressed in vessels and endothelial cells, and participates in the process of angiogenesis (Del Valle-Perez et al., 2010). Filamin B resembles filamin A (both are actin-binding proteins participating in communication between the cell membrane and the cytoskeletal network) which is associated with increased vascular permeability (Griffiths et al., 2011) and increases the risk of severe sepsis. Whether this also applies to filamin B is unclear.

End-point	SNP	Gene	Gene function	Relationship with ALL or infections	Ref.
Infectious event	rs11033797	*OR51F1*	Encodes an olfactory receptor (G-protein-coupled receptor, GPCR), intron in another gene, *MMP26*.	Not known. Theoretically functional analogy with germinal centre B-lymphocytes.	(Feldmesser et al., 2006; Lawrence et al., 2013; Otaki et al., 2005)
	rs2835265	*CBR1*	Encodes a carbonyl reductase.	Influences anthracycline metabolism.	(Bains et al., 2009)
	rs28627172	*POLDIP3*	Encodes the enzyme polymerase delta-interacting protein 3. Regulates cell growth. Associated with autoimmune disorders.	Associated with sleep patterns during dexamethasone treatment in childhood ALL patients.	(Avila et al., 2008; Vallance et al., 2010)
	rs1129844	*CCL11*	Encodes eotaxin-1, which displays chemotactic activity to eosinophils.	A possible theoretically role in host defence as an inflammatory signalling molecule.	(Simons et al., 2005)
Culture positive	rs12632456	*FLNB*	Encodes filamin B, binds actin and connects the plasma membrane to the intracellular cytoskeleton, thereby changing cell-shape. Participates in angiogenesis.	Not known. Resembles filamin A, which is associated with increased vascular permeability and is known to increase the risk of severe sepsis.	(Del Valle-Perez et al., 2010; Griffiths et al., 2011)
	rs11712186	*TOPBP1*	Encodes a binding protein that interacts with topoisomerase II beta. Catalyses transient breakage of DNA strands.	Associated with prognosis in CLL and cancer biology.	(Sellick et al., 2008)

Table II. Significant SNPs, genes and gene functions, and their relationships with ALL or infection.

Metabolic pathways related to infections during ALL treatment

Infections are considered complex diseases involving the interplay between many genes as well as host interactions with the environment. Exploring the pathways of infection seems to be more biologically relevant than performing single-gene/SNP studies. We were interested in identifying the crucial pathways involved in infections during induction therapy; this approach was possible because our starting point was the genes and SNPs with known biological functions.

We identified two pathways that were most predictive of an infectious event: "GPCR downstream signalling" and "Bile acid and salt metabolism". In addition to be the target of many drugs, many GPCRs are expressed on the cell surface of many cells, including neutrophils, were they participate in host defence and inflammation (Fig. 10) (Futosi et al., 2013).

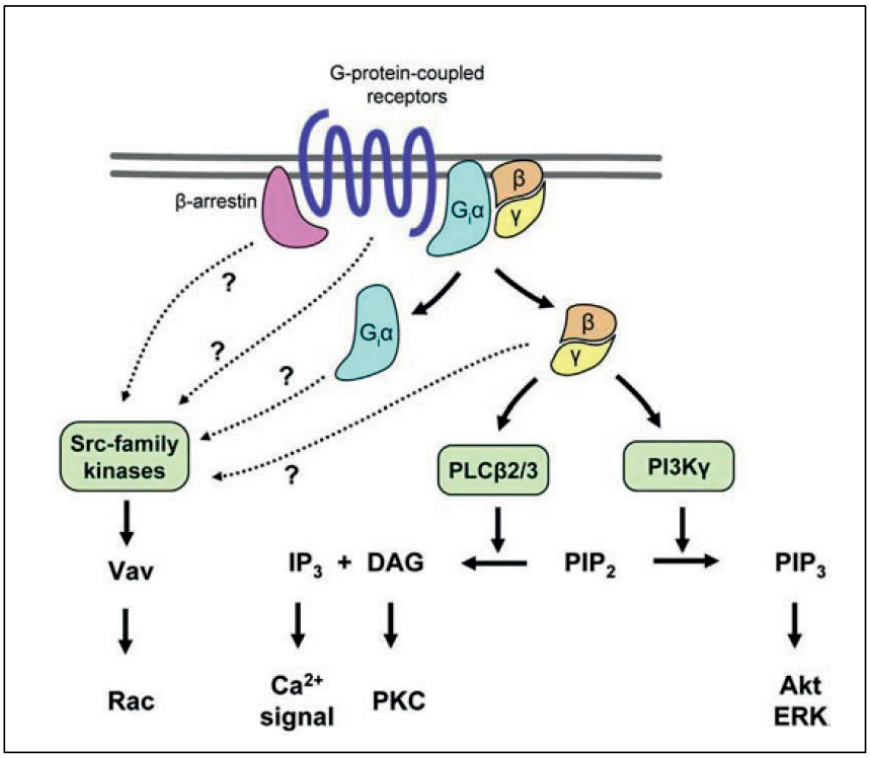


Figure 10. G-protein-coupled receptor signalling in neutrophils. G-protein-coupled receptors in neutrophils signal primarily through the Gβγ heterodimer, activating two parallel pathways through PLCβ2/3 (phospholipase Cβ2/3) and PI3Kγ (phosphatidylinositol 3Kγ). The activation of Src-family kinases likely proceeds through (an) independent and yet incompletely understood pathway(s) (question marks). Source: Futosi 2013 (Futosi et al., 2013).

There is increasing evidence that rapid effects of GCs (as opposed to delayed GC effects mediated by intracellular GC-receptors) signal via GPCRs (reviewed by Tasker et al.) (Tasker et al., 2006). One can speculate whether linkage of the "GPCR downstream signalling" pathway to infectious events in our study occurs via GPCRs in neutrophils or via GCs or both. A connection with GCs fits well because GCs are used widely during induction therapy in most childhood ALL treatment regimens including our Nordic protocols (treatment outlined in Paper I). Concerning the pathway "Bile acid and salt metabolism", some of the genes involved are also involved in GC-metabolism (*SLCO1B*-family and *CYP27A1*; Supplementary Table 4, Paper II); hence, GC-metabolism might be the common denominator. Whether these pathways also relate to deaths caused by infection is unknown but, interestingly, many of these deaths (including those caused by fungi) occurred during the first months of treatment, which overlapped with the timing of GC treatment. Again, this illustrates the immunosuppression occurring during the first part of treatment. GCs have profound effects on immune function in patients treated for inflammatory or rheumatic diseases and in childhood ALL patients (Marino et al., 2009; McDonough et al., 2008). Linking pathways relevant to infection and GC metabolism seems reasonable.

In relation to the other end-point, "culture positive", the "Interferon-signalling pathway" was the most significant pathway identified. Multiple biological functions are mediated by interferon signal transduction; e.g., anti-tumour effects, anti-viral effects, cell-cycle arrest, anti-angiogenic effects, apoptosis, cell-differentiation, immunomodulation and protein synthesis (Uddin et al., 2004). It is unclear which of these effects are relevant to immunosuppressed children with ALL, but our findings might help generating hypothesis and ideas for future research.

Archival samples

DNA extracted from archival samples, such as biopsies and smears, has great potential in genetic studies. We have evaluated the suitability of DNA extracted from archival bone marrow samples for application in multiple STR and SNP analysis. We observed significant differences between DNA concentration measured using SP and FM, and found that SP overestimated DNA concentration compared with FM. Others have also found higher concentrations of nucleotides derived from FFPE-samples measured by SP compared with FM (Deben et al., 2013). Overestimation of DNA concentration by SP analysis might be

caused by contamination with other molecules including RNA or free nucleic acids (Holden et al., 2009). Haque et al. (Haque et al., 2003) compared the DNA concentrations measured by SP with two FM methods and found that SP was the most concordant and precise DNA quantification method. Correct estimation of DNA concentration is crucial. One consequence of concentration overestimation is that the true DNA amount is less than that estimated, and there may be an insufficient starting amount of DNA for downstream analysis, which may lead to failure or bias in subsequent analyses. The most reliable DNA estimation tool in our studies seems unclear, but overestimation of DNA concentration was regarded as most hazardous; hence, the concentrations based on FM were used in our STR and SNP analysis.

Because the DNA yield varied substantially between individual archival samples and, for some samples, was very low (500 ng), we used WGA in an attempt to increase the DNA yield, especially for further downstream multiple marker analysis. We found apparently high yields of DNA after WGA when measured by SP and FM, and long fragment-lengths on gel-electrophoresis. However, WGA-samples were less effective in multiple marker analysis, especially in SNP analysis. For STR markers, we found that fragments up to 250 bp in length could be amplified and that amplification success decreased with increasing fragment length, especially for the WGA-products. Others have also had difficulties with amplification of DNA fragments >200-300 bp from FFPE samples (Talaulikar et al., 2008; Vince et al., 1998). The efficiency of WGA depends partly on DNA fragment length, and it is recommended that DNA substrates should include some fragments with lengths of at least 500 bp. This might explain, at least partly, the reduced quality of the WGA samples.

Earlier reports that included bone marrow smears from leukaemia patients focused on single markers often related to genetic aberrations in the malignant clone and, to a lesser extent, host genetics. Our findings of high detection rates in STR marker analysis based on non-WGA samples suggest that archival material might be suitable for DNA extraction and marker analysis for the study of host genetics.

We found that the detection rate was higher in the samples that had been stored for a shorter time compared with older samples. This suggests that the quality of DNA declines during longer storage periods, which is consistent with other studies (Cronin et al., 2004). However, some authors have discussed whether this relates more to the conditions of preservation, such as temperature, humidity and pH (Okello et al., 2010), whereas others have also shown that long-term storage of FFPE tissues or bone marrow slides has no significant negative effect on downstream applications (Anderegg et al., 1998; Ludyga et al., 2012). There are many possible explanations for these contradictory findings including differences in

methods for embedding, nucleic acid extraction and marker detection (Okello et al., 2010). To identify the factors most important for our material required further detailed investigations beyond the scope of this study.

Regarding multiple SNP sequencing, only a few samples could be analysed completely, and no general conclusions can be drawn. SNP profiling with multiplexed targeted sequencing is a complex laboratory procedure, and errors can be generated at many steps, not least because of poor-quality DNA extracted from archival samples. Interestingly, however, we succeeded with some samples, suggesting that multiple SNP profiling based on DNA from archival bone marrow smears might be feasible with some adjustments of the laboratory set-up. We also managed to perform SNP profiling successfully using two samples without WGA, suggesting that WGA might not be necessary but may introduce bias or noise into the analysis. However, we succeeded with five WGA samples, and we have no definitive data on which to recommend WGA. Without WGA, there might be an insufficient amount of DNA amount from one smear. A solution might be to use two or three smears per sample to increase the DNA yield from each patient, or to use other microarray-based methods, without target selection and multiplexing, that require a smaller starting amount of DNA.

Strengths and limitations

Strengths: In Paper I, data from a health registry (Nordic ALL registry in Stockholm) were used, which makes the study population-based; hence, the data on the incidence and prevalence are highly reliable, and the risk of selection bias was low. A multiple gene/SNP approach was used in which the biological functions of the genes/SNPs involved are known, making it easier to interpret the results within a biological context and making pathway analysis possible. We also demonstrated that multiple gene-association studies are possible even in small cohorts and that our results can be applied to the clinical setting. Our study design can also be applied to many other clinical end-points; for instance, to study other toxicities such as toxic death, pancreatitis, thrombosis and osteonecrosis.

Limitations: TRDs were grouped into five different groups. Grouping the causes of TRDs is problematic because there is no clear internationally accepted grouping or even a definition of TRD. Similarly, the "phenotype" used as the end-point in our SNP/gene-association study, namely "infectious event", is heterogeneous and not defined precisely, which might have contributed to bias in the associations identified. Our finding of a SNP risk profile has not

been validated in an independent patient cohort; this should be done before these findings are applied in the clinical setting.

Investigations of a set of candidate polymorphisms a priori exclude identification of other important but unknown biological mechanisms. The sample size of our cohort was small. The same findings in a larger cohort would have been more valid. DNA was available for only about 50% of patients in the original cohort, which might have contributed to selection bias.

Clinical implications

These studies highlight the importance of focusing on infections and supportive care during ALL treatment. If our SNPs/pathways are confirmed as predictive in independent cohorts or if similar SNPs/pathways are identified as risk factors for infectious TRDs, genetic risk profiling at the start of treatment may be useful for early identification of patients at risk.

In our SNP-profiling study, 41 out of 61 patients were SNP-profile positive. Of these, 40 had an event. Of the profile negative patients, 5 had an event. An intervention in the 41 profile positive patients would have given the following: i) unnecessary intervention in 1 patient, ii) lack of intervention in 5 of the event-patients with negative SNP profile, and iii) correctly intervention in 40 of 45 (89%) event-patients (Fig. 10). Subtracting the number of patients with unnecessary intervention and lack of intervention, optimized treatment would have been offered for 55 of the total of 61 (90%) patients. Such intervention may include: i) reduced or modified anti-leukaemic treatment intensity, ii) prophylactic antibiotic therapy, iii) immunological reconstitution (e.g., immunoglobulin substitution) and iv) granulocyte colony-stimulating factor therapy. Concerning reduced treatment intensity, this may only be considered for patients at risk of infectious death, and has to be balanced against a potential increase in relapse-risk.

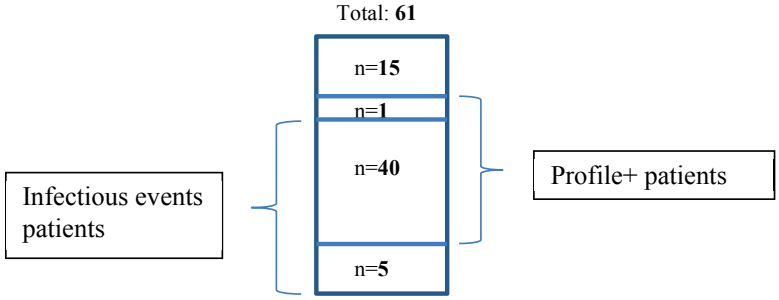


Figure 11.

The study on archival samples might not be transferred directly into the clinical setting, but is relevant for retrospective studies including study subjects who have died (TRDs). In other studies, DNA sampling can be simplified by use of archival material instead of obtaining new DNA samples (e.g., from new blood samples).

Suggestions for further research

Risk factors and causes of TRDs and other severe events are and should be a continuous focus for all cancer treatment collaborative groups. A better system to classify TRDs or childhood cancer deaths in general should be developed so that different treatment regimens can be compared more easily with regard to toxicity and survival. Our SNP risk profile including infectious TRD as the end-point, requires validation in a larger independent cohort (planned as part of our next NOPHO ALL-2014 protocol). If our findings are replicated, further biological studies including those on gene-expression and biological signalling may identify the mechanisms underlying the development of infectious events during treatment by, for example, dissecting causal genotype-phenotype relationships.

Concerning archival bone marrow samples, a more robust SNP sequencing laboratory model should be developed to include a more simplified model with fewer analysis steps in the laboratory work-flow.

During our studies on TRDs and surveillance of the ongoing NOPHO ALL-2008 protocol, we recognized high TRD rates among high-risk adolescent patients that seemed to be related to body mass index (BMI, kg/m^2). Significantly lower EFS rates among obese patients have also been reported by others (Butturini et al., 2007). Our observation initiated a new study aiming to evaluate possible relationships between BMI or BMI-changes throughout treatment and the TRD-rate in NOPHO ALL-2008 protocol patients. Preliminary results have shown a borderline significantly higher mean BMI standard deviation score among the patients who died of TRD compared with the other patients. This finding was more pronounced in the high-risk group (results presented in a poster at the NOPHO Annual Meeting in Uppsala, 2012). This case-control study is ongoing and will include height and weight data at the time of the event (TRD) to see whether changes in BMI during treatment are related to the risk of TRD.

CONCLUSIONS

In this study, we have shown that infections remain a major challenge in the treatment of childhood ALL and that infectious TRD is one of the most severe side effects of treatment. Accordingly, studies addressing the prevention of TRD have become as important as efforts to overcome the resistance to the anti-leukaemic therapy. Through our identification of a SNP-risk profile, we have shown that host genomic variation seems to play an important role in the risk of having an infectious event during treatment. Such genetic risk profiling can identify patients at risk of severe events at the start of treatment, and more risk-adapted therapy can be given with the intention of reducing severe side effects and increasing OS.

Most genetic studies rely on the use of DNA from biobanks or DNA extracted from freshly obtained samples. This presents a complication if the patient has died. We have shown that archival bone marrow samples from leukaemia patients are suitable for DNA extraction and multiple marker analysis. Although increasing the total amount of DNA, WGA reduces the analysis quality. DNA quality depends on the sample storage time, and STR-marker analysis based on archival samples of relatively short storage time (0-3 years) without WGA produces the best results. Complex SNP analysis that includes thousands of polymorphisms in a single analysis might be feasible, but the method must be optimized.

Notes

Preliminary results of this study have been presented as follows:

1) Oral presentation at NOPHO Annual Meeting in Aarhus, 2009
2) Oral presentation at NOPHO and NOBOS Annual Meeting in Tromsø, 2010
3) Poster presentation at NOPHO Annual Meeting in Uppsala, 2012
4) Poster presentation at the 54th ASH (American Society of Hematology) Annual Meeting, Atlanta, 2012

References

Afzal, S., Ethier, M. C., Dupuis, L. L., Tang, L., Punnett, A. S., Richardson, S. E., Allen, U., Abla, O., & Sung, L. (2009). Risk factors for infection-related outcomes during induction therapy for childhood acute lymphoblastic leukemia. *The Pediatric Infectious Disease Journal, 28*(12), 1064-1068.

Agbeko, R. S., Fidler, K. J., Allen, M. L., Wilson, P., Klein, N. J., & Peters, M. J. (2010). Genetic variability in complement activation modulates the systemic inflammatory response syndrome in children. *Pediatr Crit Care Med*.

Agnese, D. M., Calvano, J. E., Hahm, S. J., Coyle, S. M., Corbett, S. A., Calvano, S. E., & Lowry, S. F. (2002). Human toll-like receptor 4 mutations but not CD14 polymorphisms are associated with an increased risk of gram-negative infections. *J Infect Dis, 186*(10), 1522-1525.

Anderegg, B., Schluter, S., & Kabisch, H. (1998). Highly efficient and non-toxic DNA isolation from very old bone marrow smears. *Leukemia, 12*(5), 774-775.

Andersson, U., McKean-Cowdin, R., Hjalmars, U., & Malmer, B. (2009). Genetic variants in association studies--review of strengths and weaknesses in study design and current knowledge of impact on cancer risk. *Acta Oncologica, 48*(7), 948-954.

Arcaroli, J., Fessler, M. B., & Abraham, E. (2005). Genetic polymorphisms and sepsis. *Shock, 24*(4), 300-312.

Avila, J., Acosta, E., Machargo, M. D., Arteaga, M. F., Gallego, E., Canete, H., Garcia-Perez, J. J., & Martin-Vasallo, P. (2008). Autoantigenic nuclear proteins of a clinically atypical renal vasculitis. *J Autoimmune Dis, 5*, 3.

Bains, O. S., Karkling, M. J., Grigliatti, T. A., Reid, R. E., & Riggs, K. W. (2009). Two nonsynonymous single nucleotide polymorphisms of human carbonyl reductase 1 demonstrate reduced in vitro metabolism of daunorubicin and doxorubicin. *Drug Metab lism and Disposition: The Biological Fate of Chemicals, 37*(5), 1107-1114.

Bakhshi, S., Padmanjali, K. S., & Arya, L. S. (2008). Infections in childhood acute lymphoblastic leukemia: an analysis of 222 febrile neutropenic episodes. *Pediatr Hematol Oncol, 25*(5), 385-392.

Besag, J., & Clifford, P. (1991). Sequential Monte Carlo *p*-values. *Biometrika, 78*(2), 301-304.

Bhojwani, D., & Pui, C. H. (2013). Relapsed childhood acute lymphoblastic leukaemia. *Lancet Oncol, 14*(6), e205-217.

Blanco, E., Beyene, J., Maloney, A. M., Almeida, R., Ethier, M. C., Winick, N., Alexander, S., & Sung, L. (2012). Non-relapse mortality in pediatric acute lymphoblastic leukemia: a systematic review and meta-analysis. *Leuk Lymphoma, 53*(5), 878-885.

Blay, J. Y., Lacombe, D., Meunier, F., & Stupp, R. (2012). Personalised medicine in oncology: questions for the next 20 years. *Lancet Oncol, 13*(5), 448-449.

Boyle, E. B., Steinbuch, M., Tekautz, T., Gutman, J. R., Robison, L. L., & Perentesis, J. P. (1998). Accuracy of DNA amplification from archival hematological slides for use in genetic biomarker studies. *Cancer Epidemiol Biomarkers Prev, 7*(12), 1127-1131.

Butturini, A. M., Dorey, F. J., Lange, B. J., Henry, D. W., Gaynon, P. S., Fu, C., Franklin, J., Siegel, S. E., Seibel, N. L., Rogers, P. C., Sather, H., Trigg, M., Bleyer, W. A., & Carroll, W. L. (2007). Obesity and outcome in pediatric acute lymphoblastic leukemia. *Journal of Clinical Oncology, 25*(15), 2063-2069.

Calvano, S. E., Xiao, W., Richards, D. R., Felciano, R. M., Baker, H. V., Cho, R. J., Chen, R. O., Brownstein, B. H., Cobb, J. P., Tschoeke, S. K., Miller-Graziano, C., Moldawer,

L. L., Mindrinos, M. N., Davis, R. W., Tompkins, R. G., & Lowry, S. F. (2005). A network-based analysis of systemic inflammation in humans. *Nature, 437*(7061), 1032-1037.

Chan, J. Y., Ugrasena, D. G., Lum, D. W., Lu, Y., & Yeoh, A. E. (2011). Xenobiotic and folate pathway gene polymorphisms and risk of childhood acute lymphoblastic leukaemia in Javanese children. *Hematological Oncology, 29*(3), 116-123.

Choueiri, T. K., Je, Y., Sonpavde, G., Richards, C. J., Galsky, M. D., Nguyen, P. L., Schutz, F., Heng, D. Y., & Kaymakcalan, M. D. (2013). Incidence and risk of treatment-related mortality in cancer patients treated with the mammalian target of rapamycin inhibitors. *Annals of Oncology, 24*(8), 2092-2097.

Christensen, K., & Murray, J. C. (2007). What genome-wide association studies can do for medicine. *N Engl J Med, 356*(11), 1094-1097.

Christensen, M. S., Heyman, M., Mottonen, M., Zeller, B., Jonmundsson, G., & Hasle, H. (2005). Treatment-related death in childhood acute lymphoblastic leukaemia in the Nordic countries: 1992-2001. *Br J Haematol, 131*(1), 50-58.

Consortium, I. H. G. S. (2004). Finishing the euchromatic sequence of the human genome. *Nature, 431*(7011), 931-945.

Conter, V., Arico, M., Valsecchi, M. G., Basso, G., Biondi, A., Madon, E., Mandelli, F., Paolucci, G., Pession, A., Rizzari, C., Rondelli, R., Zanesco, L., & Masera, G. (2000). Long-term results of the Italian Association of Pediatric Hematology and Oncology (AIEOP) acute lymphoblastic leukemia studies, 1982-1995. *Leukemia, 14*(12), 2196-2204.

Cronin, M., Pho, M., Dutta, D., Stephans, J. C., Shak, S., Kiefer, M. C., Esteban, J. M., & Baker, J. B. (2004). Measurement of gene expression in archival paraffin-embedded tissues: development and performance of a 92-gene reverse transcriptase-polymerase chain reaction assay. *American Journal of Pathology, 164*(1), 35-42.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156-2158.

Davidsen, M. L., Dalhoff, K., & Schmiegelow, K. (2008). Pharmacogenetics influence treatment efficacy in childhood acute lymphoblastic leukemia. *J Pediatr Hematol Oncol, 30*(11), 831-849.

Deben, C., Zwaenepoel, K., Boeckx, C., Wouters, A., Pauwels, P., Peeters, M., Lardon, F., Baay, M., & Deschoolmeester, V. (2013). Expression analysis on archival material revisited: isolation and quantification of RNA extracted from FFPE samples. *Diagnostic Molecular Pathology, 22*(1), 59-64.

Del Valle-Perez, B., Martinez, V. G., Lacasa-Salavert, C., Figueras, A., Shapiro, S. S., Takafuta, T., Casanovas, O., Capella, G., Ventura, F., & Vinals, F. (2010). Filamin B plays a key role in vascular endothelial growth factor-induced endothelial cell motility through its interaction with Rac-1 and Vav-2. *Journal of Biological Chemistry, 285*(14), 10748-10760.

Ethier, M. C., Blanco, E., Lehrnbecher, T., & Sung, L. (2011). Lack of clarity in the definition of treatment-related mortality: pediatric acute leukemia and adult acute promyelocytic leukemia as examples. *Blood, 118*(19), 5080-5083.

Eyrich, M., Wiegering, V., Lim, A., Schrauder, A., Winkler, B., & Schlegel, P. G. (2009). Immune function in children under chemotherapy for standard risk acute lymphoblastic leukaemia - a prospective study of 20 paediatric patients. *Br J Haematol, 147*(3), 360-370.

Falconi, M., Teti, G., Zago, M., Pelotti, S., Gobbi, P., Breschi, L., & Mazzotti, G. (2007). Effect of fixative on chromatin structure and DNA detection. *Microscopy Research and Technique, 70*(7), 599-606.
Feldmesser, E., Olender, T., Khen, M., Yanai, I., Ophir, R., & Lancet, D. (2006). Widespread ectopic expression of olfactory receptor genes. *BMC Genomics, 7*, 121.
Foss, R. D., Guha-Thakurta, N., Conran, R. M., & Gutman, P. (1994). Effects of fixative and fixation time on the extraction and polymerase chain reaction amplification of RNA from paraffin-embedded tissue. Comparison of two housekeeping gene mRNA controls. *Diagnostic Molecular Pathology, 3*(3), 148-155.
Futosi, K., Fodor, S., & Mocsai, A. (2013). Neutrophil cell surface receptors and their intracellular signal transduction pathways. *Int Immunopharmacol, 17*(3), 638-650.
Ghaderi, S., Lie, R. T., Moster, D., Ruud, E., Syse, A., Wesenberg, F., & Bjorge, T. (2012). Cancer in childhood, adolescence, and young adults: a population-based study of changes in risk of cancer death during four decades in Norway. *Cancer Causes and Control, 23*(8), 1297-1305.
Gilbert, M. T., Haselkorn, T., Bunce, M., Sanchez, J. J., Lucas, S. B., Jewell, L. D., Van Marck, E., & Worobey, M. (2007). The isolation of nucleic acids from fixed, paraffin-embedded tissues-which methods are useful when? *PLoS One, 2*(6), e537.
Graubner, U. B., Porzig, S., Jorch, N., Kolb, R., Wessalowski, R., Escherich, G., & Janka, G. E. (2008). Impact of reduction of therapy on infectious complications in childhood acute lymphoblastic leukemia. *Pediatr Blood Cancer, 50*(2), 259-263.
Green, H., Paul, M., Vidal, L., & Leibovici, L. (2007). Prophylaxis for Pneumocystis pneumonia (PCP) in non-HIV immunocompromised patients. *Cochrane Database Syst Rev*(3), CD005590.
Griffiths, G. S., Grundl, M., Allen, J. S., 3rd, & Matter, M. L. (2011). R-Ras interacts with filamin a to maintain endothelial barrier function. *Journal of Cellular Physiology, 226*(9), 2287-2296.
Hann, I., Viscoli, C., Paesmans, M., Gaya, H., & Glauser, M. (1997). A comparison of outcome from febrile neutropenic episodes in children compared with adults: results from four EORTC studies. International Antimicrobial Therapy Cooperative Group (IATCG) of the European Organization for Research and Treatment of Cancer (EORTC). *Br J Haematol, 99*(3), 580-588.
Haque, K. A., Pfeiffer, R. M., Beerman, M. B., Struewing, J. P., Chanock, S. J., & Bergen, A. W. (2003). Performance of high-throughput DNA quantification methods. *BMC Biotechnol, 3*, 20.
Hargrave, D. R., Hann, II, Richards, S. M., Hill, F. G., Lilleyman, J. S., Kinsey, S., Bailey, C. C., Chessells, J. M., Mitchell, C., & Eden, O. B. (2001). Progressive reduction in treatment-related deaths in Medical Research Council childhood lymphoblastic leukaemia trials from 1980 to 1997 (UKALL VIII, X and XI). *Br J Haematol, 112*(2), 293-299.
Hill, F. G., Richards, S., Gibson, B., Hann, I., Lilleyman, J., Kinsey, S., Mitchell, C., Harrison, C. J., & Eden, O. B. (2004). Successful treatment without cranial radiotherapy of children receiving intensified chemotherapy for acute lymphoblastic leukaemia: results of the risk-stratified randomized central nervous system treatment trial MRC UKALL XI (ISRC TN 16757172). *Br J Haematol, 124*(1), 33-46.
Holden, M. J., Haynes, R. J., Rabb, S. A., Satija, N., Yang, K., & Blasic, J. R., Jr. (2009). Factors affecting quantification of total DNA by UV spectroscopy and PicoGreen fluorescence. *Journal of Agricultural and Food Chemis ry, 57*(16), 7221-7226.
Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., Down, T., Durbin, R., Eyras, E., Gilbert, J., Hammond, M., Huminiecki,

L., Kasprzyk, A., Lehvaslaiho, H., Lijnzaad, P., Melsopp, C., Mongin, E., Pettett, R., Pocock, M., Potter, S., Rust, A., Schmidt, E., Searle, S., Slater, G., Smith, J., Spooner, W., Stabenau, A., Stalker, J., Stupka, E., Ureta-Vidal, A., Vastrik, I., & Clamp, M. (2002). The Ensembl genome database project. *Nucleic Acids Research, 30*(1), 38-41.

Hudson, T. J. (2013). Genome variation and personalised cancer medicine. *Journal of Internal Medicine*.

Huttunen, R., & Aittoniemi, J. (2011). New concepts in the pathogenesis, diagnosis and treatment of bacteremia and sepsis. *J Infect, 63*(6), 407-419.

Joshi-Tope, G., Gillespie, M., Vastrik, I., D'Eustachio, P., Schmidt, E., de Bono, B., Jassal, B., Gopinath, G. R., Wu, G. R., Matthews, L., Lewis, S., Birney, E., & Stein, L. (2005). Reactome: a knowledgebase of biological pathways. *Nucleic Acids Research, 33*(Database issue), D428-432.

Kanteti, R., Yala, S., Ferguson, M. K., & Salgia, R. (2009). MET, HGF, EGFR, and PXN gene copy number in lung cancer using DNA extracts from FFPE archival samples and prognostic significance. *Journal of Environmental Pathology, Toxicology and Oncology, 28*(2), 89-98.

Karathanasis, N. V., Choumerianou, D. M., & Kalmanti, M. (2009). Gene polymorphisms in childhood ALL. *Pediatr Blood Cancer, 52*(3), 318-323.

Kawedia, J. D., Kaste, S. C., Pei, D., Panetta, J. C., Cai, X., Cheng, C., Neale, G., Howard, S. C., Evans, W. E., Pui, C. H., & Relling, M. V. (2011). Pharmacokinetic, pharmacodynamic, and pharmacogenetic determinants of osteonecrosis in children with acute lymphoblastic leukemia. *Blood, 117*(8), 2340-2347; quiz 2556.

Kidas, E., Moricke, A., Beier, R., Welte, K., Schrappe, M., Stanulla, M., & Grigull, L. (2009). Genetic polymorphisms of the lymphotoxin alpha gene are associated with increased risk for lethal infections during induction therapy for childhood acute leukemia: a case-control study. *Int J Hematol, 89*(5), 584-591.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., & Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res, 19*(9), 1639-1645.

Kaatsch, P. (2010). Epidemiology of childhood cancer. *Cancer Treatment Reviews, 36*(4), 277-285.

Lausen, B., Schmiegelow, K., Andreassen, B., Madsen, H. O., & Garred, P. (2006). Infections during induction therapy of childhood acute lymphoblastic leukemia--no association to mannose-binding lectin deficiency. *Eur J Haematol, 76*(6), 481-487.

Lausten-Thomsen, U., Madsen, H. O., & Schmiegelow, K. (2010). Characterizing chromosomal break points of t(12;21)[ETV6-RUNX1]-positive leukaemia using multiple tiling PCR on whole-genome-amplified DNA. *Leukemia, 24*(1), 203-205.

Lawrence, M. S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., Sivachenko, A., Carter, S. L., Stewart, C., Mermel, C. H., Roberts, S. A., Kiezun, A., Hammerman, P. S., McKenna, A., Drier, Y., Zou, L., Ramos, A. H., Pugh, T. J., Stransky, N., Helman, E., Kim, J., Sougnez, C., Ambrogio, L., Nickerson, E., Shefler, E., Cortes, M. L., Auclair, D., Saksena, G., Voet, D., Noble, M., DiCara, D., Lin, P., Lichtenstein, L., Heiman, D. I., Fennell, T., Imielinski, M., Hernandez, B., Hodis, E., Baca, S., Dulak, A. M., Lohr, J., Landau, D. A., Wu, C. J., Melendez-Zajgla, J., Hidalgo-Miranda, A., Koren, A., McCarroll, S. A., Mora, J., Lee, R. S., Crompton, B., Onofrio, R., Parkin, M., Winckler, W., Ardlie, K., Gabriel, S. B., Roberts, C. W., Biegel, J. A., Stegmaier, K., Bass, A. J., Garraway, L. A., Meyerson, M., Golub, T. R., Gordenin, D. A., Sunyaev, S., Lander, E. S., & Getz, G. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature, 499*(7457), 214-218.

Lehrnbecher, T., Bernig, T., Hanisch, M., Koehl, U., Behl, M., Reinhardt, D., Creutzig, U., Klingebiel, T., Chanock, S. J., & Schwabe, D. (2005). Common genetic variants in the interleukin-6 and chitotriosidase genes are associated with the risk for serious infection in children undergoing therapy for acute myeloid leukemia. *Leukemia, 19*(10), 1745-1750.

Lehrnbecher, T., Varwig, D., Kaiser, J., Reinhardt, D., Klingebiel, T., & Creutzig, U. (2004). Infectious complications in pediatric acute myeloid leukemia: analysis of the prospective multi-institutional clinical trial AML-BFM 93. *Leukemia, 18*(1), 72-77.

Levinsen, M., Shabaneh, D., Bohnstedt, C., Harila-Saari, A., Jonsson, O. G., Kanerva, J., Lindblom, A., Lund, B., Andersen, E. W., & Schmiegelow, K. (2012). Pneumocystis jiroveci pneumonia prophylaxis during maintenance therapy influences methotrexate/6-mercaptopurine dosing but not event-free survival for childhood acute lymphoblastic leukemia. *European Journal of Haematology, 88*(1), 78-86.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics, 25*(14), 1754-1760.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics, 26*(5), 589-595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079.

Ludyga, N., Grunwald, B., Azimzadeh, O., Englert, S., Hofler, H., Tapio, S., & Aubele, M. (2012). Nucleic acids from long-term preserved FFPE tissues are suitable for downstream analyses. *Virchows Archiv, 460*(2), 131-140.

Marino, S., Verzegnassi, F., Tamaro, P., Stocco, G., Bartoli, F., Decorti, G., & Rabusin, M. (2009). Response to glucocorticoids and toxicity in childhood acute lymphoblastic leukemia: role of polymorphisms of genes involved in glucocorticoid response. *Pediatr Blood Cancer, 53*(6), 984-991.

McDonough, A. K., Curtis, J. R., & Saag, K. G. (2008). The epidemiology of glucocorticoid-associated adverse events. *Current Opinion in Rheumatology, 20*(2), 131-137.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res, 20*(9), 1297-1303.

McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P., & Cunningham, F. (2010). Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics, 26*(16), 2069-2070.

Middleton, P. G., Cullup, H., Dickinson, A. M., Norden, J., Jackson, G. H., Taylor, P. R., & Cavet, J. (2002). Vitamin D receptor gene polymorphism associates with graft-versus-host disease and survival in HLA-matched sibling allogeneic bone marrow transplantation. *Bone Marrow Transplant, 30*(4), 223-228.

Mira, J. P., Cariou, A., Grall, F., Delclaux, C., Losser, M. R., Heshmati, F., Cheval, C., Monchi, M., Teboul, J. L., Riche, F., Leleu, G., Arbibe, L., Mignon, A., Delpech, M., & Dhainaut, J. F. (1999). Association of TNF2, a TNF-alpha promoter polymorphism, with septic shock susceptibility and mortality: a multicenter study. *JAMA, 282*(6), 561-568.

Mullighan, C. G. (2012). Molecular genetics of B-precursor acute lymphoblastic leukemia. *Journal of Clinical Investigation, 122*(10), 3407-3415.

Neth, O., Hann, I., Turner, M. W., & Klein, N. J. (2001). Deficiency of mannose-binding lectin and burden of infection in children with malignancy: a prospective study. *Lancet, 358*(9282), 614-618.

Nguyen, K., Devidas, M., Cheng, S. C., La, M., Raetz, E. A., Carroll, W. L., Winick, N. J., Hunger, S. P., Gaynon, P. S., & Loh, M. L. (2008). Factors influencing survival after relapse from acute lymphoblastic leukemia: a Children's Oncology Group study. *Leukemia, 22*(12), 2142-2150.

NOPHO Annual Report 2013. (2013).

Okello, J. B., Zurek, J., Devault, A. M., Kuch, M., Okwi, A. L., Sewankambo, N. K., Bimenya, G. S., Poinar, D., & Poinar, H. N. (2010). Comparison of methods in the recovery of nucleic acids from archival formalin-fixed paraffin-embedded autopsy tissues. *Analytical Biochemistry, 400*(1), 110-117.

Otaki, J. M., Fearon, D. T., & Yamamoto, H. (2005). The proto-oncogene BCL-6 is expressed in olfactory sensory neurons. *Neuroscience Research, 53*(2), 189-200.

Ottenhoff, T. H., Verreck, F. A., Lichtenauer-Kaligis, E. G., Hoeve, M. A., Sanal, O., & van Dissel, J. T. (2002). Genetics, cytokines and human infectious disease: lessons from weakly pathogenic mycobacteria and salmonellae. *Nature Genetics, 32*(1), 97-105.

Pabst, T., Schwaller, J., Tobler, A., & Fey, M. F. (1996). Detection of microsatellite markers in leukaemia using DNA from archival bone marrow smears. *British Journal of Haematology, 95*(1), 135-137.

Picard. [Internet] Available from: http://picard.sourceforge.net.

Pichler, H., Reismuller, B., Steiner, M., Dworzak, M. N., Potschger, U., Urban, C., Meister, B., Schmitt, K., Panzer-Grumayer, R., Haas, O. A., Attarbaschi, A., & Mann, G. (2013). The inferior prognosis of adolescents with acute lymphoblastic leukaemia (ALL) is caused by a higher rate of treatment-related mortality and not an increased relapse rate--a population-based analysis of 25 years of the Austrian ALL-BFM (Berlin-Frankfurt-Munster) Study Group. *British Journal of Haematology, 161*(4), 556-565.

Prucker, C., Attarbaschi, A., Peters, C., Dworzak, M. N., Potschger, U., Urban, C., Fink, F. M., Meister, B., Schmitt, K., Haas, O. A., Gadner, H., & Mann, G. (2009). Induction death and treatment-related mortality in first remission of children with acute lymphoblastic leukemia: a population-based analysis of the Austrian Berlin-Frankfurt-Munster study group. *Leukemia, 23*(7), 1264-1269.

Pui, C. H. (2006). *Childhood leukemias* (Second edition ed.). New York: Cambridge University Press.

Pui, C. H., Mahmoud, H. H., Wiley, J. M., Woods, G. M., Leverger, G., Camitta, B., Hastings, C., Blaney, S. M., Relling, M. V., & Reaman, G. H. (2001). Recombinant urate oxidase for the prophylaxis or treatment of hyperuricemia in patients With leukemia or lymphoma. *J Clin Oncol, 19*(3), 697-704.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J., & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics, 81*(3), 559-575.

Rahiala, J., Perkkio, M., & Riikonen, P. (1998). Infections occurring during the courses of anticancer chemotherapy in children with ALL: a retrospective analysis of 59 patients. *Pediatr Hematol Oncol, 15*(2), 165-174.

Rubnitz, J. E., Lensing, S., Zhou, Y., Sandlund, J. T., Razzouk, B. I., Ribeiro, R. C., & Pui, C. H. (2004). Death during induction therapy and first remission of acute leukemia in childhood: the St. Jude experience. *Cancer, 101*(7), 1677-1684.

Rungoe, C., Malchau, E. L., Larsen, L. N., & Schroeder, H. (2010). Infections during induction therapy for children with acute lymphoblastic leukemia. the role of sulfamethoxazole-trimethoprim (SMX-TMP) prophylaxis. *Pediatr Blood Cancer, 55*(2), 304-308.

Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., Sherry, S., Mullikin, J. C., Mortimore, B. J., Willey, D. L., Hunt, S. E., Cole, C. G., Coggill, P. C., Rice, C. M., Ning, Z., Rogers, J., Bentley, D. R., Kwok, P. Y., Mardis, E. R., Yeh, R. T., Schultz, B., Cook, L., Davenport, R., Dante, M., Fulton, L., Hillier, L., Waterston, R. H., McPherson, J. D., Gilman, B., Schaffner, S., Van Etten, W. J., Reich, D., Higgins, J., Daly, M. J., Blumenstiel, B., Baldwin, J., Stange-Thomann, N., Zody, M. C., Linton, L., Lander, E. S., & Altshuler, D. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature, 409*(6822), 928-933.

Schmiegelow, K., Forestier, E., Hellebostad, M., Heyman, M., Kristinsson, J., Soderhall, S., & Taskinen, M. (2010). Long-term results of NOPHO ALL-92 and ALL-2000 studies of childhood acute lymphoblastic leukemia. *Leukemia, 24*(2), 345-354.

Schmiegelow, K., Forestier, E., Kristinsson, J., Soderhall, S., Vettenranta, K., Weinshilboum, R., & Wesenberg, F. (2009). Thiopurine methyltransferase activity is related to the risk of relapse of childhood acute lymphoblastic leukemia: results from the NOPHO ALL-92 study. *Leukemia, 23*(3), 557-564.

Schaaf, B. M., Boehmke, F., Esnaashari, H., Seitzer, U., Kothe, H., Maass, M., Zabel, P., & Dalhoff, K. (2003). Pneumococcal septic shock is associated with the interleukin-10-1082 gene promoter polymorphism. *Am J Respir Crit Care Med, 168*(4), 476-480.

Sellick, G. S., Wade, R., Richards, S., Oscier, D. G., Catovsky, D., & Houlston, R. S. (2008). Scan of 977 nonsynonymous SNPs in CLL4 trial patients for the identification of genetic variants influencing prognosis. *Blood, 111*(3), 1625-1633.

Shamim, Z., Ryder, L. P., Heilmann, C., Madsen, H., Lauersen, H., Andersen, P. K., Svejgaard, A., Jacobsen, N., & Muller, K. (2006). Genetic polymorphisms in the genes encoding human interleukin-7 receptor-alpha: prognostic significance in allogeneic stem cell transplantation. *Bone Marrow Transplant, 37*(5), 485-491.

Simons, J. E., Rothenberg, M. E., & Lawrence, R. A. (2005). Eotaxin-1-regulated eosinophils have a critical role in innate immunity against experimental Brugia malayi infection. *European Journal of Immunology, 35*(1), 189-197.

Slats, A. M., Egeler, R. M., van der Does-van den Berg, A., Korbijn, C., Hahlen, K., Kamps, W. A., Veerman, A. J., & Zwaan, C. M. (2005). Causes of death--other than progressive leukemia--in childhood acute lymphoblastic (ALL) and myeloid leukemia (AML): the Dutch Childhood Oncology Group experience. *Leukemia, 19*(4), 537-544.

Spits, C., Le Caignec, C., De Rycke, M., Van Haute, L., Van Steirteghem, A., Liebaers, I., & Sermon, K. (2006). Whole-genome multiple displacement amplification from single cells. *Nat Protoc, 1*(4), 1965-1970.

Srinivasan, M., Sedmak, D., & Jewell, S. (2002). Effect of fixatives and tissue processing on the content and integrity of nucleic acids. *American Journal of Pathology, 161*(6), 1961-1971.

Stabell, N., Nordal, E., Stensvold, E., Gammelsrud, K. W., Lund, B., Taxt, A., Buhring, F., Greve-Isdahl, M., Fornebo, H. P., Simonsen, G. S., & Klingenberg, C. (2008). Febrile neutropenia in children with cancer: a retrospective Norwegian multicentre study of clinical and microbiological outcome. *Scand J Infect Dis, 40*(4), 301-307.

Stanulla, M., Schrappe, M., Brechlin, A. M., Zimmermann, M., & Welte, K. (2000). Polymorphisms within glutathione S-transferase genes (GSTM1, GSTT1, GSTP1) and risk of relapse in childhood B-cell precursor acute lymphoblastic leukemia: a case-control study. *Blood, 95*(4), 1222-1228.

Statistics Norway [Internet] Available from: https://www.ssb.no.

Sutherland, A. M., & Walley, K. R. (2009). Bench-to-bedside review: Association of genetic variation with sepsis. *Crit Care, 13*(2), 210.

Talaulikar, D., Gray, J. X., Shadbolt, B., McNiven, M., & Dahlstrom, J. E. (2008). A comparative study of the quality of DNA obtained from fresh frozen and formalin-fixed decalcified paraffin-embedded bone marrow trephine biopsy specimens using two different methods. *Journal of Clinical Pathology, 61*(1), 119-123.

Tasker, J. G., Di, S., & Malcher-Lopes, R. (2006). Minireview: rapid glucocorticoid signaling via membrane-associated receptors. *Endocrinology, 147*(12), 5549-5556.

Therneau, T. M., & Atkinson, E. J. (1997). An introduction to recursive partitioning using the rpart routines. *Technical Report* (Vol. 61): Mayo Clinic, Rochester. http://www.mayo.edu/hsr/techrpt/61.pdf.

Thompson, E. R., Herbert, S. C., Forrest, S. M., & Campbell, I. G. (2005). Whole genome SNP arrays using DNA derived from formalin-fixed, paraffin-embedded ovarian tumor tissue. *Hum Mutat, 26*(4), 384-389.

Toft, N., Birgens, H., Abrahamsson, J., Bernell, P., Griskevicius, L., Hallbook, H., Heyman, M., Holm, M. S., Hulegardh, E., Klausen, T. W., Marquart, H. V., Jonsson, O. G., Nielsen, O. J., Quist-Paulsen, P., Taskinen, M., Vaitkeviciene, G., Vettenranta, K., Asberg, A., & Schmiegelow, K. (2013). Risk group assignment differs for children and adults 1-45 yr with acute lymphoblastic leukemia treated by the NOPHO ALL-2008 protocol. *European Journal of Haematology*.

U.S. Food and Drug Administration [Internet] Available from: http://www.fda.gov.

Uddin, S., & Platanias, L. C. (2004). Mechanisms of type-I interferon signal transduction. *J Biochem Mol Biol, 37*(6), 635-641.

Vallance, K., Liu, W., Mandrell, B. N., Panetta, J. C., Gattuso, J. S., Hockenberry, M., Zupanec, S., Yang, L., Yang, J., & Hinds, P. S. (2010). Mechanisms of dexamethasone-induced disturbed sleep and fatigue in paediatric patients receiving treatment for ALL. *European Journal of Cancer, 46*(10), 1848-1855.

van de Wetering, M. D., de Witte, M. A., Kremer, L. C., Offringa, M., Scholten, R. J., & Caron, H. N. (2005). Efficacy of oral prophylactic antibiotics in neutropenic afebrile oncology patients: a systematic review of randomised controlled trials. *Eur J Cancer, 41*(10), 1372-1382.

Vince, A., Poljak, M., & Seme, K. (1998). DNA extraction from archival Giemsa-stained bone-marrow slides: comparison of six rapid methods. *Br J Haematol, 101*(2), 349-351.

Wesolowska, A., Dalgaard, M. D., Borst, L., Gautier, L., Bak, M., Weinhold, N., Nielsen, B. F., Helt, L. R., Audouze, K., Nersting, J., Tommerup, N., Brunak, S., Sicheritz-Ponten, T., Leffers, H., Schmiegelow, K., & Gupta, R. (2011). Cost-effective multiplexing before capture allows screening of 25 000 clinically relevant SNPs in childhood acute lymphoblastic leukemia. *Leukemia, 25*(6), 1001-1006.

Wheeler, K., Chessells, J. M., Bailey, C. C., & Richards, S. M. (1996). Treatment related deaths during induction and in first remission in acute lymphoblastic leukaemia: MRC UKALL X. *Arch Dis Child, 74*(2), 101-107.

Yang, J. J., Cheng, C., Yang, W., Pei, D., Cao, X., Fan, Y., Pounds, S. B., Neale, G., Trevino, L. R., French, D., Campana, D., Downing, J. R., Evans, W. E., Pui, C. H., Devidas, M., Bowman, W. P., Camitta, B. M., Willman, C. L., Davies, S. M., Borowitz, M. J., Carroll, W. L., Hunger, S. P., & Relling, M. V. (2009). Genome-wide interrogation of germline genetic variation associated with treatment response in childhood acute lymphoblastic leukemia. *JAMA: the journal of the American Medical Association, 301*(4), 393-403.

Yildirim, I., Aytac, S., Ceyhan, M., Cetin, M., Tuncer, M., Cengiz, A. B., Secmeer, G., & Yetgin, S. (2008). Piperacillin/tazobactam plus amikacin versus carbapenem

monotherapy as empirical treatment of febrile neutropenia in childhood hematological malignancies. *Pediatric Hematology and Oncology, 25*(4), 291-299.

Aaltonen, K. E., Ebbesson, A., Wigerup, C., & Hedenfalk, I. (2011). Laser capture microdissection (LCM) and whole genome amplification (WGA) of DNA from normal breast tissue --- optimization for genome wide array analyses. *BMC Res Notes, 4*, 69.

Appendix

TREATMENT RELATED DEATHS IN NOPHO ALL-2000

Patient Identification Number:
(Set by study coordinator =NOPHOnr)

1. **Date of birth: _____(ddmmyyyy)** Gender: Female □
Male □

Immunoglobulines at diagnosis:

a. a.IgA _____
b. b.IgM _____
c. c.IgG _____
Units in: micromol/l □ gram/l □

Viral immune status before treatment (set "x"):

	Specific IgG			PCR		
	pos	**neg**	**not done**	**pos**	**Neg**	**not done**
CMV						
EBV						
HBV						
HCV						
HIV						
Parvovirus B19						
HSV						
Morbilli						
VZV						

ALL therapy overall:

a. Death before any antileukemic therapy: Yes □ No □

Clinical features before start of treatment:	**Yes**	**No**	**Unknown**	**If "yes" <u>duration</u> from start of symptom to start of treatment:**
Fever (>37,5°C ax)				
Infection				
Respiratory tract symptoms				
Bleeding				
Petechiae				
Skin rash				
Fatigue				
Headache				
Limb/joint pain				
Other symptoms (specify):				

b. Date of start: _____ Treatment protocol: _____
c. Major changes from protocol treatment:
   i. Yes □ No □
      If so, specify: _____
      _____
d. BMT:
   Yes□ No□
   1. Conditioning regimen: _____
   2. Conditioning start date: _____
   3. Transplantation date: _____

Recent ALL– therapy:

Specify all chemotherapy given during **the last 3 weeks** before death (or send copy of chemotherapy forms)

Drug	Total dose pr m²	Date begin (ddmmyyyy)	Date end (ddmmyyyy)

Clinical features and physical examination at start of the last course of chemotherapy before death:

Start date: _____

Blood samples at start: White bloodcell count: _____ $x10^9$/L ; Neutrophil count: _____$x10^9$/L; Lymphocyte count: _____$x10^9$/L; Thrombocyte count:_____$x10^9$/L; Hemoglobin: _____ g/L; mmol/L

	Yes	No	Unknown	
Fever (>37,5°C ax)				
Infection				Local □ Systemic □
Respiratory distress				
Bleeding				If yes, site:
Skin rash				If yes, site:
Splenomegaly				Cm below costal margin:
Hepatomegaly				Cm below costal margin:
Lymphadenopathy ≥1cm				If yes, site:
CNS symptoms/signs				If yes, specify:
Other symptoms (specify):				

2. Complications the last 3 weeks preceding death:

	Yes	No	Unknown	Organism(s):
Pneumonia	□	□	□	_____
Bact. sepsis	□	□	□	_____
Septic shock	□	□	□	_____
Meningitis	□	□	□	_____
Viral infection	□	□	□	_____
				Site:_____
Fungal inf.	□	□	□	_____
				Site:_____
Artificial ventil.	□	□	□	Start date: _____ End date:_____
Surgery	□	□	□	Which: _____
				Date:_____
Bleeding	□	□	□	
Encephalopathy	□	□	□	

Hepatic toxicity/failure (ULN=Upper Limit Normal; LLN=Lower Limit Normal):

* ALAT: Normal □ >ULN-2,5 x ULN □
  >2,5-5,0 x ULN □ >5,0 x ULN □
* Bilirubin: Normal □ >ULN-1,5 x ULN □
  >1,5-3,0 x ULN □ >3,0 x ULN □
* Albumin: Normal □ <LLN-3,0 g/dL □
  ≥2,0-3,0 g/dL □ <2,0 g/dL □
* Other signs of liver failure:

  _____

Renal toxicity/failure :

* Creatinin: Normal □ ULN-1,5 x ULN □
  >1,5-3,0 x ULN □ >3,0 x ULN □
* Requiring dialysis? Yes□ No□
* Other signs of renal desease/failure:

  _____

Cardiac failure:

* Left ventricular function:
  - Normal □
  - Fraction of shortening (FS) ≥24% - <30% □
  - FS <24% □
  - Congestive Heart failure □

Other, specify: _____

_____

Clinical features and physical examination at the time of death:

	Yes	No	Unknown	
Fever (>37,5°C ax)				
Infection				Local □ Systemic □ If local, site:
Respiratory distress				
Bleeding				If yes, site:
Skin rash				If yes, site:
Splenomegaly				Cm below costal margin:
Hepatomegaly				Cm below costal margin:
Lymphadenopathy ≥1cm				If yes, site:
CNS symptoms/signs				If yes, specify:
Other symptoms (specify):				

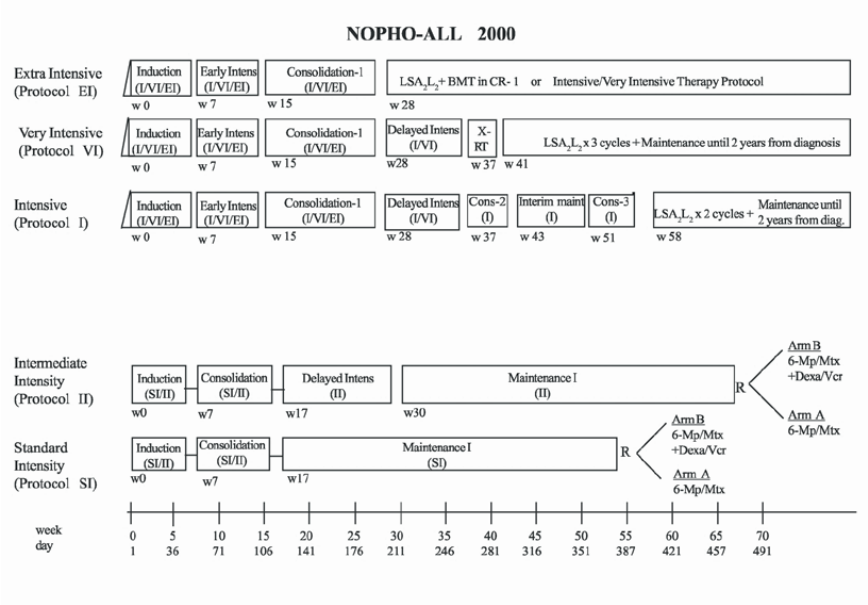
Artificial implants present during the last 3 weeks before death:

a. IV access: Port-a-cath □ No implants □
Hickman-catheter □
Other(specify): _____

b. Other artificial implants: _____

Date of death: _____

* Set a mark in the protocol for the time of death:



Hematological data before death:

Last bone marrow examination before death: Date: _____

Percentage of blasts: **M0** □ **M1** □ **M2** □ **M3** □

Date							
Days before death	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**	**0**
WBC (x10^9/L):							
ANC (x10^9/L):							
Hemoglobin (g/L;mmol/L):							
Thrombocyte count (x10^9/L):							
Lymphocyte count (x10^9/L):							
CRP							

Max CRP during ongoing infection: _____ Date: _____

Microbiological analysis during the last 3 weeks before death:

Date	**Material**	**Test (culture, PCR, etc)**	**Organism**

Antibiotics during the last 3 weeks before death :

(All kinds included, even PCP prophylaxis)

Start date	**Stop date**	**Antibiotics**	**Reason for change (set "x"):**		
			Specific organism /resistance pattern	**Resistant bacteria**	**Lack of clinical response**

Cause of death:

a) Autopsy performed: Yes□ No□

b) Main cause of death:

Tumor lysis compl.	□	Interstitial pneumonitits	□
Leukostasis	□	ARDS	□
Thrombosis	□	Cardiac toxicity	□
Bleeding	□	Encephalopathy	□
Renal failure	□	Multi organ failure	□
VOD	□	Unknown	□

Infection:

Viral □ Organism: _____
Site:_____

Bacterial □ Organism: _____
Site:_____

Fungal □ Organism: _____
Site: _____

Parasitic □ Organism: _____
Site: _____

Unknown □

Other, specify: _____

c) Contributing causes of death: _____

_____

Further comments:

_____

_____

_____

_____

_____

_____

Trondheim, Norway, January, 2008

On behalf of the NOPHO ALL-event group,

Bendik Lund, MD (study coordinator) **Ann Elisabeth Åsberg, MD**

Kjeld Schmiegelow, Professor Dr. Med.

Papers I-III

Paper I

Pediatr Blood Cancer 2011;56:551–559

Risk Factors for Treatment Related Mortality in Childhood Acute Lymphoblastic Leukaemia

Bendik Lund, MD,[1,2,*] Ann Åsberg, MD, PhD,[1] Mats Heyman, MD, PhD,[3] Jukka Kanerva, MD, PhD,[4] Arja Harila-Saari, MD, PhD,[5] Henrik Hasle, MD, PhD,[6] Stefan Söderhäll, MD, PhD,[3] Ólafur Gisli Jónsson, MD,[7] Stian Lydersen, PhD,[8] and Kjeld Schmiegelow, MD, PhD[2,9] On behalf of the Nordic Society of Paediatric Haematology and Oncology (NOPHO)

Background. In spite of major improvements in the cure rate of childhood acute lymphoblastic leukaemia (ALL), 2–4% of patients still die from treatment related complications. ***Procedure.*** We investigated the pattern of treatment related deaths (TRDs) and possible risk factors in the NOPHO ALL-92 and ALL-2000 protocols. Fifty-five TRDs were identified among the 1,645 ALL-92 patients and 33 among the 1,090 ALL-2000 patients. ***Results.*** There was no significant difference in the incidence of TRDs between the two protocols (3.4% vs. 3.2%). Five patients died before initiation of therapy (0.2%), and the overall subsequent risk of induction death and death in first complete remission (CR1) was 1.2% and 1.8%, respectively. Infections were the major cause of death comprising 72% of all cases including 9 deaths from *Pseudomonas aeruginosa* and 11 deaths from fungal infections. Other causes of death included bleeding or thrombosis (eight patients), tumour burden related toxicities (seven patients) and organ toxicity (seven patients). Female gender (hazard ratio (HR): 2.2, 95% confidence interval (95% CI): 1.4–3.4), high white blood cell count ($\geq 200 \times 10^9$/L) at diagnosis (HR: 3.5, 95% CI: 1.7–7.1), T-cell disease (HR: 1.9, 95% CI: 1.01–3.7), Down syndrome (HR: 7.3, 95% CI: 3.6–14.9) and haematopoietic stem cell transplantation in CR1 (HR: 8.0, 95% CI: 3.3–19.5) were identified as independent risk factors for TRD. ***Conclusion.*** Several TRDs were potentially preventable and future efforts should be directed towards patients at risk. Pediatr Blood Cancer. 2011;56:551–559. © 2010 Wiley-Liss, Inc.

Key words: acute lymphoblastic leukaemia; paediatric oncology; risk factors; toxicity; treatment-related death

INTRODUCTION

Improved risk grouping and intensification of chemotherapy have significantly reduced the relapse rate of childhood acute lymphoblastic leukaemia (ALL) [1–3]. In contrast, and in spite of improved supportive care, treatment related deaths (TRDs) continue to occur in 2–4% of the patients (Table I) [4–11]. Thus, the relative significance of TRDs among all events has increased because of the decreasing relapse-rate in current treatment protocols. TRD represents the 'tip of the iceberg' of the total toxicity related to modern treatment of childhood ALL.

Infections, bleeding or thrombosis, tumour burden complications, and therapy induced organ toxicities are the most common causes of TRD [4,7,12,13]. Four major factors influence the risk of these and other severe, although non-fatal, toxicities: the leukaemia itself (e.g., the tumour burden and specific organ involvement), the treatment intensity, the supportive care (including specific guidelines and physician and patient compliance to these) and host factors (including inherited genetic polymorphisms that influence drug disposition and immune function) [2,14–16].

To identify potentially preventable risk factors for specific TRDs, we explored all 88 TRDs among 2,735 ALL patients treated on two consecutive Nordic protocols from 1992 to 2008.

MATERIALS AND METHODS

Since 1992 all children with ALL in the five Nordic countries (Denmark, Finland, Iceland, Norway and Sweden) have been treated according to common Nordic protocols [1]. Long-term results have recently been published for the NOPHO ALL-92 study showing a 10-year event-free survival (EFS) of $74.6 \pm 1.1\%$ and an overall survival of $84.7 \pm 0.9\%$. For the NOPHO ALL-2000 protocol, the 5-year EFS was $79.4 \pm 1.5\%$ and the overall survival was $89.1 \pm 1.1\%$ [17]. Between January 1992 and June 2008, 2,882 children 1.0–14.9 years of age with B-cell precursor or T-cell ALL were diagnosed within the Nordic countries. We excluded the following patients from this study: 2 Down syndrome patients who received no anti-leukaemic therapy (1 with significant co-morbidity and 1 diagnosed post-mortem), 14 patients not treated according to NOPHO ALL-92 or ALL-2000 protocols, 41 patients treated according to the NOPHO ALL-2000 protocol before it was officially opened, 20 patients treated according to the international protocol for Ph+ (t(9;22)(q34;q11)/*BCR-ABL* fusion) ALL, 14 patients who were

DOI 10.1002/pbc.22719
Published online 8 December 2010 in Wiley Online Library (wileyonlinelibrary.com).

[1]Department of Pediatrics, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway; [2]Department of Pediatrics, The University Hospital Rigshospitalet, Copenhagen, Denmark; [3]Department of Pediatrics, Astrid Lindgrens Barnsjukhus, Stockholm, Sweden; [4]Hospital for Children and Adolescents, University of Helsinki, Helsinki, Finland; [5]Department of Paediatrics and Adolescence, Oulu University Hospital, Oulu, Finland; [6]Department of Pediatrics, Aarhus University Hospital, Skejby, Denmark; [7]Department of Pediatrics, University Hospital, Reykjavik, Iceland; [8]Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology (NTNU), Trondheim, Norway; [9]The Faculty of Medicine, Institute of Gynaecology, Obstetrics and Paediatrics, University of Copenhagen, Copenhagen, Denmark

Grant sponsor: The Danish Childhood Cancer Foundation; Grant sponsor: Childrens Cancer Foundation, Department of Pediatrics, St. Olavs Hospital, Trondheim University Hospital; Grant sponsor: The Swedish Childhood Cancer Foundation.

Conflict of interest: Nothing to declare.

*Correspondence to: Bendik Lund, Department of Pediatrics, St. Olavs Hospital HF, Trondheim University Hospital, N-7006, Trondheim, Norway. E-mail: bendik.lund@ntnu.no

Received 10 January 2010; Accepted 25 May 2010

TABLE I. Treatment Related Death (TRD) in Childhood Acute Lymphoblastic Leukaemia (ALL) and Study Groups

Study group	Years of study	Age range (years)	No. of patients included	Total no of TRDs, no. of pat. (%)	Pre-treatment deaths, [a] no. of pat. (%)	Induction deaths, no. of pat. (%)	Death in CR1, no. of pat. (%)	Death post-HSCT in CR1, no. of pat. (%)
Present study	1992–2008	1–15	2,735	88 (3.2)	5 (0.2)	34 (1.2)	49 (1.8)	10 (0.4)
Moricke et al. [6], BFM-95	1995–2000	0–18	2,169	62 (2.9)	Excluded	16 (0.7)	33 (1.5)	13 (0.6)
Hargrave et al. [9], MRC UKALL XI	1991–1997	1–15	2,090	56 (2.7)	Excluded	25 (1.2)	27 (1.3)	4 (0.2)
Vilmer et al. [8], CLCG-EORTC 58881	1989–1998	0–18	2,065	76 (3.7)	10 (0.5)	9 (0.4)	57 (2.8)	Not reported
Conter et al. [5], AIEOP-91 *Study 91*	1991–1995	0–15	1,194	[b](3.2)	Excluded	[b](1.4)	[b](1.8)	Not reported
Rubnitz et al. [7], SJCRH	1984–1999	0–18	1,011	36 (3.6)	Not reported	14 (1.4)	16 (1.7)	6 (0.6)
Slats et al. [11], DCOG	1984–1996	0–15	875	29 (3.3)	3 (0.3)	6 (0.7)	14 (1.6)	5 (0.7)
Prucker et al. [4], A-BFM	1981–1999	0–21	896	31 (3.5)	Excluded	7 (0.8)	24 (2.7)	Excluded
Moghrabi et al. [10], DFCI 95-01	1996–2000	0–18	491	7 (1.4)	Not reported	4 (0.8)	3 (0.6)	Not reported

AIEOP, Associazione Italiana Ematologia Oncologia Pediatrica; BFM, Berlin-Frankfurt-Münster Study Group; SJCRH, St. Jude Children's Research Hospital; CLCG-EORTC, Children Leukemia Cooperative Group—European Organisation for Research and Treatment of Cancer; DFCI, The Dana-Farber Cancer institute; MRC UKALL, Medical Research Council United Kingdom Acute Lymphoblastic Leukaemia; NOPHO, Nordic Society of Paediatric Haematology and Oncology; DCOG, Dutch Childhood Oncology Group; A-BFM, Austrian Berlin-Frankfurt-Münster Study Group; CR1, first complete remission. Included is death post-HSCT when reported; HSCT, haematopoietic stem cell transplantation. [a]Death before start of treatment; [b]Numbers not available.

diagnosed and started on anti-leukaemic therapy outside the Nordic countries, 52 patients who changed protocol during therapy (from ALL-92 to ALL-2000, or from ALL-2000 to ALL-2008), 1 patient who died from Leighs encephalopathy day 6 during induction therapy and three patients with bilineage leukaemia. Thus, 2,735 patients were included in the present study (Fig. 1). Patients with resistant disease to NOPHO treatment were included until the day resistant disease was recognised. Patients dying from a second malignant neoplasm or after relapse were not considered as TRD in this study. Since the purpose of this study was to explore the risk of TRD among children 1.0–14.9 years of age in order to improve the treatment and supportive care for non-infants and non-Ph + ALL patients, we excluded patients treated according to other protocols, including the international Interfant and EsPhALL protocols. Of the 1,231 females and 1,504 males, 1,645 patients were treated according to the NOPHO ALL-92 protocol (open from January 1992 until October 2001), and 1,090 patients were treated according to the NOPHO ALL-2000 protocol (open from January 2002 until June 2008). There were 2,431 (89%) B-cell precursor and 277 (10%) T-cell ALL patients (no information on lineage was available for 27 patients). Fifty-nine had Down syndrome and 68 patients were registered with CNS leukaemia at diagnosis. Nine hundred and fifty-five (35%) were standard risk, 942 (34%) intermediate risk and 838 (31%) were high-risk patients according to the ALL-92 and ALL-2000 criteria. Patients who stayed in first complete remission (CR1) were followed until April 2009. Data were obtained from the prospective registration of all patients in the NOPHO leukaemia registry. Additional clinical data concerning cause of death from the patient files for the 88 TRDs were obtained through a questionnaire sent out to the treating centres. Fifty-five of the 88 TRDs in this study have previously been published [12].

Risk Grouping

Patients were stratified into three risk groups: standard risk (SR), intermediate risk (IR) and high-risk (HR) ALL. In this study, SR and IR ALL are combined as low-risk ALL. Stratification criteria for the low-risk groups were (all criteria needed): WBC $<50.0 \times 10^9$/L, B-cell precursor ALL, no CNS or testicular leukaemia, no unfavourable cytogenetic alterations (i.e. 11q23/*MLL*-rearrangements, t(9;22)(q34;q11)/*BCR-ABL* fusion, t(1;19)(q23;p13)/*E2A-PBX1* fusion, hypodiploidy (<45 chromosomes)), and good response to initial therapy defined as M1 or M2 bone marrow at day 15 and M1 bone marrow at day 29. The HR patients fulfilled at least one of the high-risk criteria listed above. In the ALL-2000 protocol, HR patients with at least one of the following criteria were allocated to haematopoietic stem cell transplantation (HSCT) in CR1: M3 bone marrow on day 29, 11q23/*MLL* rearrangements and age $\geq$10 years, t(9;22)(q34;q11)/*BCR-ABL* fusion, a karyotype with a modal chromosome number <34, or initial WBC $\geq 200 \times 10^9$/L. There were no uniform Nordic criteria for HSCT in the ALL-92 protocol.

Induction Remission Treatment, Both Protocols

The first 50 days of therapy included only small differences between the two protocols and consisted of the following: induction included prednisolone (60 mg/m^2/day) day 1–36 followed by 10 days tapering. Pre-treatment for 1–6 days with prednisolone dose increments was used in cases with large leukaemic cell burden at

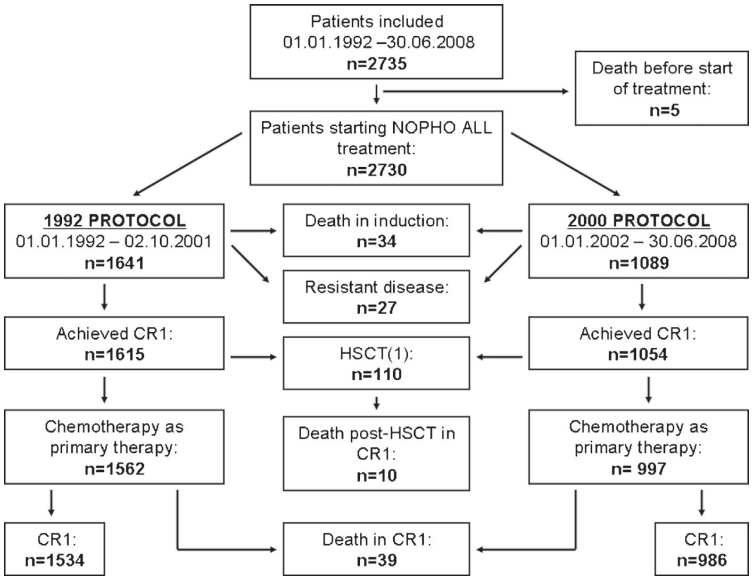


Fig. 1. Flow chart of patients included in the study.

diagnosis to decrease the risk of tumour lysis syndrome (TLS). Vincristine (Vcr) was given weekly for 6 weeks in doses of 2.0 mg/m^2; the maximum dose was increased from 2.0 mg in the ALL-92 protocol to 2.5 mg in the ALL-2000 protocol. In the ALL-92 protocol doxorubicin at doses of 40 mg/m^2 was given three and four times in the low-risk and high-risk groups, respectively. In the ALL-2000 protocol the corresponding number was reduced to two or three doses for the low- and high-risk groups, respectively. Methotrexate (MTX) was given intrathecally (i.t.) at days 1, 8, 15 and 29 in age-adjusted doses. For the ALL-92 patients Erwinia asparaginase was used (dose: 30,000 IU/m^2 daily on days 37–46), and for the ALL-2000 patients, *Escherichia coli* asparaginase was used (dose: 6,500 IU/m^2 at 3- to 4-day intervals up to a total of four doses starting treatment day 36).

Further Treatment

NOPHO ALL-92 protocol: Detailed information concerning therapy for the NOPHO ALL-92 protocol has been published earlier [1,18]. Treatment duration from the day of diagnosis was 2.5 years for the SR group and 2.0 years for the other groups. There were no specific guidelines for supportive care in the ALL-92 protocol.

NOPHO ALL-2000 protocol: Following induction, the low-risk groups received identical consolidation therapy consisting of 6-MP (25 mg/m^2/day) and alternating blocks with high-dose MTX (5 g/m^2/24 hr with i.t. MTX and Leucovorin rescue) and low-dose cytarabine (75 mg/m^2/day for 4 days, two times). After induction, early intensification therapy followed for the high-risk groups, which included two doses of cyclophosphamide (1,000 mg/m^2) 4 weeks apart with low-dose cytarabine (75 mg/m^2 daily for two 4 days periods), oral 6-thioguanine (6-TG) and two doses of i.t. MTX. Consolidation therapy for the high-risk groups included alternating courses of high-dose MTX (8 g/m^2/24 hr with i.t. MTX and Leucovorin rescue) and high-dose cytarabine (12 g/m^2) times two or four, with two 2 months intervening periods of oral weekly MTX and daily 6-MP with two Vcr/prednisolone reinductions per period. In the low-risk groups the interval between high-dose MTX courses was increased from 2 to 4 weeks compared to the ALL-92 protocol, and the start of Leucovorin rescue was delayed 6 hours to "hour 42" from start of the MTX infusion. Patients with IR- or HR-ALL received delayed intensification therapy with dexamethasone (doses: IR patients 6 mg/m^2/day, HR patients 10 mg/m^2/day) for 2 weeks, weekly Vcr (2.0 mg/m^2) for 4 weeks, weekly doxorubicin (HR) or daunorubicin (IR), at a dose of 30 mg/m^2/day, three (HR) or four (IR) times. In addition, *E. coli* asparaginase (dose: 6,500 IU/m^2) was given four times followed by cyclophosphamide (dose: 1,000 mg/m^2), low-dose cytarabine and 6-TG. Maintenance therapy consisted of weekly oral MTX (starting dose: 20 mg/m^2) and daily oral 6-MP, dose adjusted according to TPMT activity (starting doses, wild-type patients: 75 mg/m^2/day, heterozygous patients: 50 mg/m^2/day, TPMT deficient patients: 5–10 mg/m^2/day). In addition, low-risk patients received alternate pulses at four-week intervals Vcr (2.0 mg/m^2, one dose)/dexamethasone (6 mg/m^2/day for 5 days) and high-dose MTX (5 g/m^2/24 hr) times five during the first year of maintenance therapy. HR patients received reinductions every four weeks throughout maintenance therapy consisting of Vcr (2.0 mg/m^2) and dexamethasone (6 mg/m^2/day for 5 days). After 1 year in maintenance patients in the low-risk groups were randomised into two arms: one arm with only 6-MP/MTX, and one arm with 6-MP/MTX plus an additional eight pulses every 6 weeks of dexamethasone (dose: orally 6 mg/m^2/day) and Vcr (dose: i.v. 2.0 mg/m^2, max. dose 2.5 mg) for 5 days. For HR patients who were not treated with HSCT in CR1, two cycles of the LSA_2L_2 regimen [19] was inserted at the beginning of maintenance therapy. Children above 5 years of age with T-cell ALL and mediastinal mass, and/or WBC at diagnosis $100 < 200 \times 10^9$/L, and/or CNS leukaemia at diagnosis

received cranial irradiation (24 or 18 Gy depending on whether or not CNS leukaemia was present at diagnosis). The treatment duration was 2.5 years for low-risk patients and 2.0 years for high-risk patients.

The ALL-2000 protocol included guidelines for supportive care concerning TLS, hyperleukocytosis, superior vena cava syndrome and superior mediastinal syndrome. There were no general recommendations for prevention of infection with *Pneumocystis jiroveci* (PJ), fungal infections or management of febrile neutropenia. The use of myeloid growth factors was optional after courses with high-dose cytarabine (2 g/m^2).

Causes of Death

The causes of death were grouped as (i) related to tumour burden (i.e. TLS or leukostasis with compromised organ function due to infiltrating blast cells), (ii) bleeding or thrombosis, (iii) infections, (iv) therapy-induced organ toxicity and (v) other or uncertain cause. The criteria for infectious deaths were clinical signs of infections in combination with fever and/or raised C-reactive protein and/or microbiologically proven infection, and no other obvious cause of death. An autopsy was performed in 34 (39%) cases.

Statistical Analysis and Definitions

Proportions were compared by Chi-square tests. Kaplan–Meier plots and survival tables were used for survival analysis including estimation of cumulative incidence of TRD and subgroups were compared using Log-rank tests. The main event in the analysis was TRD including (i) pre-treatment death (death before any anti-leukaemic therapy), (ii) induction death (death after start of treatment, but before achieved remission) and (iii) death in CR1 (included deaths happening up to 6 months after end of treatment). Patients who experienced other events (resistant disease, relapse or second malignant neoplasms) were censored at the time of these events. Time to TRD was defined as time between diagnosis of ALL and date of death. Patients who experienced no events were censored at the time of last follow-up. Cox proportional hazard regression analysis was performed with time to TRD (or time to infectious TRD) as only event, and the following covariates were included: sex, age ($<$ or $\geq$10 years), WBC ($<$ or $\geq 200 \times 10^9$/L), B-cell precursor versus T-cell disease, protocol (ALL-92 vs. ALL-2000), presence or absence of CNS-disease, Down syndrome and HSCT in CR1. Age was dichotomised at 10 years since this was used as stratification criteria for the intermediate risk group in both protocols. For patients who underwent HSCT in CR1, a time-dependent covariate was defined as zero before, and one after the date of HSCT, and for all other patients, this time-dependent covariate was defined as zero. We checked for the possible 10 two-way interactions among the 5 significant covariates, using a stepwise approach with Bonferroni adjusted significance level $0.05/10 = 0.005$. In all other analyses, P-values <0.05 were considered significant. All tests were two-sided. The statistical analyses were performed using the SPSS 16.0 statistical software.

Ethical Considerations

The protocols were approved by the national or regional ethics committees in the five Nordic countries, and the study was conducted in accordance with the Declaration of Helsinki.

RESULTS

Of the total of 2,735 patients, 51 females and 37 males (3.2%) with a median age of 4 years (75%, range 2.0–11.0 years) at diagnosis experienced a TRD. The TRDs constituted 25% of all 354 deaths in the study population with 240 deaths after relapse as the largest group. According to risk groups (stratification based on up-front criteria) the cumulative incidence of TRD was $1.7 \pm 0.4\%$, $2.4 \pm 0.5\%$ and $6.7 \pm 0.9\%$ for the SR, IR and HR group, respectively ($P < 0.001$). Of all patients, five deaths (0.2%) occurred before initiation of treatment (pre-treatment deaths), 34 deaths (1.2%) occurred during remission induction and 49 deaths (1.8%) happened during post-induction treatment in CR1 including 10 deaths after HSCT (Fig. 1). Of the 88 TRDs, 63 (72%) patients died from infections, 8 (9%) died from bleeding or thrombosis, 7 (8%) died from organ toxicity and 7 (8%) died from tumour burden complications. In addition, two HSCT patients died from severe graft-versus-host disease (GVHD) of which one of them had additional respiratory failure, and one patient died from an erroneous procedure (intrathecal injection of Vcr). The cumulative incidence of TRD in the ALL-92 and ALL-2000 protocol did not differ significantly, and was $3.4 \pm 0.5\%$ and $3.2 \pm 0.6\%$, respectively ($P = 0.85$).

Tumour Burden Related Early Deaths

Of the seven patients (six boys) who died from tumour burden related problems (of which five died pre-treatment), six patients had a WBC $\geq 350 \times 10^9$/L at diagnosis. All, except one patient, died from intracerebral infiltration of leukaemic blast cells with or without intracerebral bleeding. One patient with a large mediastinal tumour (WBC at diagnosis: 23×10^9/L) was resuscitated because of cardiac arrest during anaesthesia for diagnostic bone marrow aspiration and died after 11 days due to secondary brain damage. No patients died from TLS complications.

Treatment Related Death in Relation to Time and Phase of Therapy

The TRDs occurred at a median of 6 weeks from diagnosis (Fig. 2) and 76% of the non-HSCT related deaths occurred within the first 80 days of treatment. The annual proportion of TRD before remission (pre-treatment deaths and induction deaths) ranged from 0.0–2.7% (calendar years 1992–2007). Of the 2,730 patients who started anti-leukaemic treatment, 34 patients died during remission induction (Fig. 1) yielding a proportion of induction death of 1.2% (0.7% in the low-risk groups and 2.4% in the high-risk group, $P < 0.001$). The causes of induction deaths were infections (n = 26), tumour burden (n = 2) and bleeding or thrombosis (n = 6).

Of the 2,669 patients who achieved remission, 49 died in CR1 (including 10 post-HSCT patients) giving a proportion of death in CR1 of 1.8% (1.3% in the low-risk groups and 3.2% in the high-risk group, $P = 0.001$). Of these, 17 died during the last part of the induction phase, 3 died during early intensification, 2 during consolidation, 3 during late intensification, 12 during maintenance (none of which during the LSA_2L_2 regimen) and 1 during cranial irradiation. For one patient, exact information concerning treatment phase was lacking. Of the 12 patients who died during maintenance, only 2 infectious TRDs occurred within 6 weeks from administration of pulses of Vcr/dexamethasone. Of the 39 non-HSCT TRDs, 30 (77%) patients died from infections and 6 (15%) from organ toxicity includ-

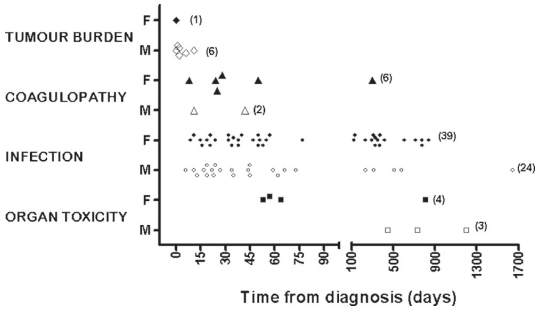


Fig. 2. Causes of death in relation to time from diagnosis for 85 out of 88 treatment related deaths (TRDs) in the NOPHO ALL-92 and NOPHO ALL-2000 protocol (TRDs post-HSCT included). Not shown are one patient who died from an accidental intrathecal injection of vincristine, and two post-haematopoietic stem cell transplantation patients who died from severe GVHD. F: Female (n = 1,231). M: Male (n = 1,504).

ing two deaths from acute pancreatitis after the second and third dose of asparaginase, respectively. Other deaths from organ toxicity included one MTX-related and one hypertensive encephalopathy, one toxic hepatitis following a high-dose MTX course and one patient who died shortly after cessation of therapy of haemolytic uremic syndrome of unknown origin. One patient died 6 weeks after diagnosis from an intestinal bleeding, one Down syndrome patient died after 10 months from an intracerebral bleeding and one patient died after an erroneously administered intrathecal dose of Vcr.

Of the 110 patients who underwent HSCT in CR1, 10 patients died from treatment related complications; 5 died from infections, 1 from leukoencephalopathy and 2 from GVHD. In two HSCT patients who died, CMV infection was suspected, but not proven.

Infectious Deaths

Of the 63 infectious deaths (Table III), 7 cases were polybacterial or polymicrobial, and in 16 cases no microorganism was found. Chemotherapy induced neutropenia (defined as neutrophils $\leq 0.5 \times 10^9$/L) during the last week of life was found in relation to 29 (48%) of the infectious deaths.

Bleeding and Thrombosis

Of the eight TRDs caused by bleeding or thrombosis (leukostasis-associated TRDs excluded), six deaths occurred within the first 50 days from diagnosis including four patients who died before the first dose of asparaginase. Of the bleeding deaths, two of the haemorrhages were located to the brain, one to the intestines and one haemorrhage was diffuse involving multiple organs. Of the four who died from thrombosis/infarction, three occurred in the brain and one in the intestines.

Risk Factors

In simple Cox regression analyses several closely associated clinical features were linked to an increased risk of TRDs: T-cell disease, WBC $\geq 200 \times 10^9$/L at diagnosis, presence of CNS disease and HSCT in CR1 (Table II). Kaplan–Meier plot for WBC is shown in Figure 3. In multiple Cox regression analysis, female gender, WBC $\geq 200 \times 10^9$/L at diagnosis, T-cell disease, Down syndrome, and HSCT in CR1 were identified as independent risk factors for

TABLE II. Risk Factors for Treatment Related Death (TRD), NOPHO ALL-1992 and NOPHO ALL-2000 Protocol

Risk factors	TRD (n = 88)	All patients (n = 2735)	HR (95% CI), simple regression	Adjusted HR (95% CI), multiple regression
Sex				
Female	51	1,231	1.7 (1.1–2.6)	2.2 (1.4–3.4)
Male	37	1,504	1.0	1.0
Age (years)				
1–9	71	2,274	1.0	1.0
10–14	17	461	1.2 (0.7–2.1)	0.9 (0.51–1.5)
WBC ($\times 10^9$/L)				
<200	69	2,603	1.0	1.0
≥200	19	132	6.5 (3.9–10.8)	3.5 (1.7–7.1)
Immunophenotype				
T-cell	22	277	3.2 (2.0–5.3)	1.9 (1.01–3.7)
Not T-cell	66	2,458	1.0	1.0
Protocol				
1992	55	1,645	1.0 (0.7–1.6)	1.2 (0.76–1.8)
2000	33	1,090	1.0	1.0
Down syndrome				
Yes	9	59	5.6 (2.8–11.2)	7.3 (3.6–14.9)
No	79	2,676	1.0	1.0
CNS				
Yes	7	68	3.8 (1.7–8.2)	2.1 (0.93–4.6)
No	81	2667	1.0	1.0
HSCT in CR1				
Yes	10	109	17.7 (8.1–38.6)	8.0 (3.3–19.5)
No	78	2,626	1.0	1.0

HR = hazard ratio; 95% CI = 95% confidence interval; WBC = white blood cell count at diagnosis; HSCT in CR1 = haematopoietic stem cell transplantation in first complete remission. A time-dependent covariate was constructed for HSCT patients (see text).

TABLE III. Causative Organisms in the 63 Infectious Deaths in the NOPHO ALL-1992 and ALL-2000 Protocol

		Number of patients	
Major group	Organism	ALL-1992	ALL-2000
Bacteria	*Coagulase negative staphylococcus*	1	
	Bacillus cereus[a]		1
	Pseudomonas aeruginosa[†]	8	1
	Eschericia coli	3	1
	Klebsiella	1	
	Listeria[a]		1
	Stenotrophomonas (Xanthomonas) *maltophilia*	1	
	Enterobacter	1	
	Stomatococcus	1	
	Micrococcus	1	
	Unspecified		2
Virus	*Cytomegalovirus***	1	2
	Adenovirus	1	
	Influenza B	1	
	Respiratory syncytial virus		1
Fungi	*Candida**	2	3
	Geotrichum capitatum		1
	Pneumocystis jirovechi	1	
	Unspecified[†]	3	1
Polybacterial or microbial**,[†]	6	1	
Unspecified microorganism**,[††]	9	7	
Total	41	22	

The number of daggers indicate the number of Down syndrome cases (n = 5) in respective groups. The number of asterisks indicate the number of deaths in respective groups post-haematopoietic stem cell transplantation (HSCT) (n = 7). [a]Central nervous system infections (n = 2).

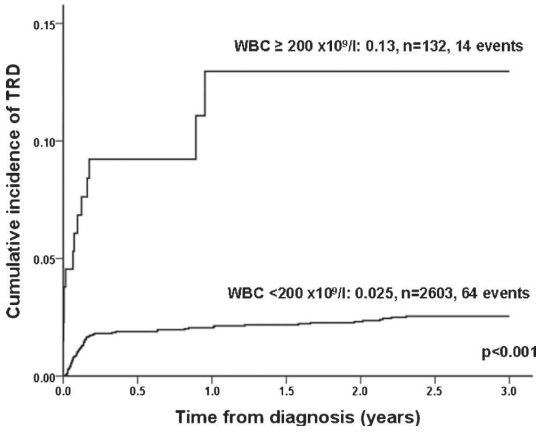


Fig. 3. Cumulative incidence of treatment related death (TRD) by white blood cell count at diagnosis (WBC). Patients with WBC $<200 \times 10^9$/L ($2.5 \pm 0.3\%$ (standard error)) versus patients with WBC $\geq 200 \times 10^9$/L ($13.0 \pm 3.6\%$) at diagnosis (Kaplan–Meier estimator, $P < 0.001$, Log-rank test). Patients who underwent HSCT in CR1 were censored at the time of HSCT.

TRD (Table II). The only significant interaction was Down syndrome and WBC $\geq 200 \times 10^9$/L ($P < 0.001$), where the presence of both these factors attenuated the hazard. However, this did not significantly affect the effects of the other risk factors in Table II. The haemoglobin level at diagnosis was not found to be a risk factor for TRD. Analysing infectious TRDs separately, females more often died from infections compared to males (adjusted HR: 2.4, 95% CI: 1.4–4.0). Of the nine Down syndrome TRDs, seven occurred before achieved remission.

DISCUSSION

Despite a steady improvement in supportive care over the last 30 years, treatment related toxicity remains a major challenge in childhood ALL therapy, and the risk of treatment related mortality has not decreased substantially in the last 20 years [4,7,12,20]. Furthermore, in parallel to the decreasing number of patients dying from the leukaemia itself (i.e. tumour burden related death, resistant disease and relapse), TRDs comprise an increasing proportion of the overall mortality (25% in our study). In addition, TRD represents the most severe form of overall treatment related toxicity, and preventive efforts directed towards factors influencing TRD will also affect toxicity in general.

The TRDs can be subdivided into infections, tumour burden complications, organ toxicity and bleeding/thrombosis with infections as the most frequent cause. To reduce toxicity and prevent TRD, interventions are needed to address the risk factors, that is the tumour burden, the treatment intensity and the supportive care. In addition, there is a need of increased knowledge of relevant genetic host factors.

Concerning the seven deaths from tumour burden related problems within the first days from diagnosis, the outcome for patients like these is clearly related to the rapidity of the diagnostic work-up and initiation of therapy. Traditionally, patients with a high tumour burden have been treated with intravenous alkaline fluids, a xanthine oxidase inhibitor (allopurinol), and gradually increasing doses of a corticosteroid, delaying more intensive chemotherapy until the blast

count has fallen and thereby lowering the risk of TLS. Modern treatment with recombinant urate oxidase (rasburicase) has been shown to be safe and effective for prophylaxis or treatment of the urate related problems of TLS in childhood malignancies [21]. Rasburicase produces a rapid decrease in plasma uric acid concentration and makes it possible to start with tumour reducing therapy within hours [21,22]. Since a mononuclear white blood cell count above 200 is virtually pathognomonic for leukaemia, rapid initiation of full dose corticosteroid therapy should be considered after administration of urate oxidase at the presentation of such patients and sampling of peripheral blood for the diagnostic leukaemia work-up. This approach could potentially prevent the tumour burden related TRDs seen in this study. In case of significant electrolyte disturbances, for example hyperphosphatemia, these should be corrected before start of anti-leukaemic therapy.

The treatment-induced immunosuppression includes neutropenia, impaired humoral antibody response, impaired cell-mediated immunity, phagocytic defects and disturbed cytokine function [23,24]. Furthermore, Eyrich et al. [25] recently showed that B-cells were most severely affected throughout therapy and did not recover before the end of therapy. T-cells and natural killer cells partially recovered at the end of induction therapy and are the dominating lymphocyte subset during maintenance therapy. We have shown that infections remain the most common cause of TRD comprising 72% of all cases, which is in line with the findings of most other groups [4,9,13,26].

It is noteworthy that 76% of all TRDs occurred during the first 80 days of treatment when the immune deficiency is most pronounced due to the tumour load itself (i.e. infiltrating blast cells in bone marrow and other lymphoid tissue) and very intensive chemotherapy including corticosteroids.

Of the patients dying from infections, 48% had chemotherapy-induced neutropenia during the week preceding death. However, the choice of empirical anti-microbial therapy for febrile neutropenia varies widely between treating centres even within the same collaborative group. In a British survey of 21 United Kingdom Children's Cancer Study Group (UKCCSG) centres treating children with febrile neutropenia, the management varied both concerning the definition of fever, the definition of neutropenia, and in the choice of empirical antibiotic therapy [27]. Studies have shown that combination therapy of a *Pseudomonas*-covering beta-lactam (e.g. ceftazidime) and an aminoglycoside is superior to monotherapy in case of *Pseudomonas* infections [28,29]. Out of the nine *Pseudomonas* TRDs in our study, only one occurred in the ALL-2000 protocol. A likely explanation for this is an increased use of empiric *Pseudomonas*-covering antibiotic therapy in case of febrile neutropenia, but no data on routine supportive care have been registered as part of this study.

There was no uniform approach to PJ infection prophylaxis during the study period and no patient-specific data on the use of PJ-prophylaxis is available. Thus, it is unknown to what extent local strategies for PJ-prophylaxis have contributed to the low frequency of mortality from PJ (Table III).

Concerning antibacterial prophylaxis, there is no uniform approach in the Nordic region. The use of prophylactic TMP-SMX has earlier been shown to reduce the incidence of both PCP infections and other infections and bacteraemia in ALL patients [30,31]. A retrospective Danish study comparing two different patients groups, one receiving TMP-SMX prophylaxis during induction treatment, and one without, showed that the TMP-SMX group had significantly fewer episodes of fever and fewer fever-related positive blood cultures during induction therapy [32]. In a review article on efficacy of oral prophylactic antibiotics in neutropenic afebrile oncology patients (including 22 clinical trials) van de Wetering et al. [33] concluded that oral prophylactic antibiotics decreased Gram-negative bacteraemia and infection related mortality. However, only 3 of the 22 reviewed trials included paediatric patients. Larger prospective studies are needed to explore if antibacterial prophylaxis during the neutropenic phases of the first 3 months of anti-leukaemic therapy can reduce the risk of TRDs in childhood ALL.

Of the 11 deaths due to fungal infections, 9 occurred within the first 9 weeks of treatment. Early death from fungal infections has also been found by others [7], pointing at the potential advantage of anti-fungal antibiotic prophylaxis at least during the early part of treatment. One possible disadvantage when using azoles as anti-fungal prophylaxis in combination with weekly administration of Vcr (as during induction therapy) is increased Vcr-related neurotoxicity. Vcr is metabolised by CYP3A enzymes and azoles are potent inhibitors of CYP3A isoenzymes resulting in higher Vcr concentrations [34,35]. Fluconazole is a relatively weaker inhibitor of CYP3A compared to other azoles (e.g. itraconazole) [35] leaving fluconazole as a reasonable drug of choice, although it has a limited effect on invasive aspergillosis and fluconazole-resistant candida strains [36,37]. In the British UKCCSG study on febrile neutropenia, the strategy for empirical anti-fungal treatment was not described in detail [27], and further studies are needed to address this issue.

Some of the patients in our study died of infections despite seemingly adequate antibiotic treatment according to the resistance pattern of the microorganism in question. This raises the question of the role of critical host factors including pharmacogenetics and immunogenetics. Studies on genetic polymorphisms in immunoregulating mediators have shown an association with outcome during childhood leukaemia induction therapy [14], in childhood AML patients [38], in childhood malignancies in general [16] and in post-HSCT patients [39,40]. However, most studies are candidate gene based involving only a few genes, and genome-wide studies of variations within the immune response involving multiple genes and haplotypes are lacking. Identification of possible immunogenetic risk profiles at the start of therapy could potentially be helpful for risk adapted and individualised supportive care. Possible preventive measures for patients at significantly increased risk of infectious TRD could include: (i) reduced or modified anti-leukaemic therapy intensity, (ii) prophylactic antibiotic therapy, (iii) immunological reconstitution (e.g. immunoglobuline substitution) and (iv) granulocyte colony-stimulating factor therapy.

Subgroups of patients dying from specific organ toxicities (bleeding or thrombosis included) constitute a relatively small fraction of TRDs. The two deaths from asparaginase induced acute pancreatitis are rare events and most patients experiencing acute pancreatitis during treatment are successfully treated [41]. Another frequent complication from asparaginase treatment is thrombotic events occurring in up to 36% of patients [42]. In our study, the only patient possibly dying from an asparaginase related thrombotic event was a Down syndrome patient who died on treatment day 50 from a thrombosis in the right common carotid artery resulting in massive cerebral infarction. Of the four patients who died from bleeding complications (tumour burden associated bleedings excluded), three had accompanying severe thrombocytopenia (platelets $<20 \times 10^9$/L). Three of the patients also had an ongoing

infection, illustrating the additive risk of infection and coagulation disturbances. Very large, probably international, genetic studies such as those performed by the Ponte di Legno group (see Biondi et al. [43]) are needed to identify patients at excessive risk of very rare fatal events such as MTX-induced encephalopathy and liver failure.

In conclusion, TRD remains a major challenge and constitutes an increasing fraction of all deaths in childhood ALL. Accordingly, studies addressing the prevention of TRDs have become as important as efforts to overcome the resistance to the anti-leukaemic therapy.

ACKNOWLEDGMENTS

This study has received financial support from The Danish Childhood Cancer Foundation; Childrens Cancer Foundation, Department of Pediatrics, St. Olavs Hospital, Trondheim University Hospital; and The Swedish Childhood Cancer Foundation. Some results from this study (data from the NOPHO ALL-1992 protocol) have partly been published earlier [12]. Kjeld Schmiegelow holds the Danish Childhood Cancer Foundation professorship in Paediatric Oncology. *Authorship*: Bendik Lund has written the paper and, together with Kjeld Schmiegelow, designed the study which was based on the design of the study by Christensen et al. [12]. It has been written in close collaboration with the NOPHO Event group where Jukka Kanerva is the present chair. All co-authors have participated in the writing process and interpretation of the data. Mats Heyman has compiled and scrutinised the data in the NOPHO leukaemia registry. Ann Åsberg, Arja Harila-Saari, Henrik Hasle, Stefan Söderhäll (members of NOPHO Event-group) and Ólafur Gisli Jónsson provided clinical data through their roles as national coordinators for this study. Statistical analyses were performed by Bendik Lund and Stian Lydersen.

REFERENCES

1. Gustafsson G, Schmiegelow K, Forestier E, et al. Improving outcome through two decades in childhood ALL in the Nordic countries: The impact of high-dose methotrexate in the reduction of CNS irradiation. Nordic Society of Pediatric Haematology and Oncology (NOPHO). Leukemia 2000;14:2267–2275.
2. Pui CH, Evans WE. Treatment of acute lymphoblastic leukemia. N Engl J Med 2006;354:166–178.
3. Stanulla M, Schrappe M. Treatment of childhood acute lymphoblastic leukemia. Semin Hematol 2009;46:52–63.
4. Prucker C, Attarbaschi A, Peters C, et al. Induction death and treatment-related mortality in first remission of children with acute lymphoblastic leukemia: A population-based analysis of the Austrian Berlin-Frankfurt-Munster study group. Leukemia 2009;23:1264–1269.
5. Conter V, Arico M, Valsecchi MG, et al. Long-term results of the Italian Association of Pediatric Hematology and Oncology (AIEOP) acute lymphoblastic leukemia studies, 1982–1995. Leukemia 2000;14:2196–2204.
6. Moricke A, Reiter A, Zimmermann M, et al. Risk-adjusted therapy of acute lymphoblastic leukemia can decrease treatment burden and improve survival: Treatment results of 2169 unselected pediatric and adolescent patients enrolled in the trial ALL-BFM 95. Blood 2008;111:4477–4489.
7. Rubnitz JE, Lensing S, Zhou Y, et al. Death during induction therapy and first remission of acute leukemia in childhood: The St. Jude experience. Cancer 2004;101:1677–1684.
8. Vilmer E, Suciu S, Ferster A, et al. Long-term results of three randomized trials (58831, 58832, 58881) in childhood acute lymphoblastic leukemia: A CLCG-EORTC report. Children Leukemia Cooperative Group. Leukemia 2000;14:2257–2266.
9. Hargrave DR, Hann II, Richards SM, et al. Progressive reduction in treatment-related deaths in Medical Research Council childhood lymphoblastic leukaemia trials from 1980 to 1997 (UKALL VIII, X and XI). Br J Haematol 2001;112:293–299.
10. Moghrabi A, Levy DE, Asselin B, et al. Results of the Dana-Farber Cancer Institute ALL Consortium Protocol 95-01 for children with acute lymphoblastic leukemia. Blood 2007;109:896–904.
11. Slats AM, Egeler RM, van der Does-van den Berg A, et al. Causes of death—other than progressive leukemia—in childhood acute lymphoblastic (ALL) and myeloid leukemia (AML): The Dutch Childhood Oncology Group experience. Leukemia 2005;19:537–544.
12. Christensen MS, Heyman M, Mottonen M, et al. Treatment-related death in childhood acute lymphoblastic leukaemia in the Nordic countries: 1992–2001. Br J Haematol 2005;131:50–58.
13. Wheeler K, Chessells JM, Bailey CC, et al. Treatment related deaths during induction and in first remission in acute lymphoblastic leukaemia: MRC UKALL X. Arch Dis Child 1996;74:101–107.
14. Kidas E, Moricke A, Beier R, et al. Genetic polymorphisms of the lymphotoxin alpha gene are associated with increased risk for lethal infections during induction therapy for childhood acute leukemia: A case-control study. Int J Hematol 2009;89:584–591.
15. Davidsen ML, Dalhoff K, Schmiegelow K. Pharmacogenetics influence treatment efficacy in childhood acute lymphoblastic leukemia. J Pediatr Hematol Oncol 2008;30:831–849.
16. Neth O, Hann I, Turner MW, et al. Deficiency of mannose-binding lectin and burden of infection in children with malignancy: A prospective study. Lancet 2001;358:614–618.
17. Schmiegelow K, Forestier E, Hellebostad M, et al. Long-term results of NOPHO ALL-92 and ALL-2000 studies of childhood acute lymphoblastic leukemia. Leukemia 2010;24:345–354.
18. Schmiegelow K, Al-Modhwahi I, Andersen MK, et al. Methotrexate/6-mercaptopurine maintenance therapy influences the risk of a second malignant neoplasm after childhood acute lymphoblastic leukemia: Results from the NOPHO ALL-92 study. Blood 2009;113:6077–6084.
19. Wollner N, Exelby PR, Lieberman PH. Non-Hodgkin's lymphoma in children: A progress report on the original patients treated with the LSA2-L2 protocol. Cancer 1979;44:1990–1999.
20. Hill FG, Richards S, Gibson B, et al. Successful treatment without cranial radiotherapy of children receiving intensified chemotherapy for acute lymphoblastic leukaemia: Results of the risk-stratified randomized central nervous system treatment trial MRC UKALL XI (ISRC TN 16757172). Br J Haematol 2004;124:33–46.
21. Pui CH, Mahmoud HH, Wiley JM, et al. Recombinant urate oxidase for the prophylaxis or treatment of hyperuricemia in patients With leukemia or lymphoma. J Clin Oncol 2001;19:697–704.
22. Coiffier B, Altman A, Pui CH, et al. Guidelines for the management of pediatric and adult tumor lysis syndrome: An evidence-based review. J Clin Oncol 2008;26:2767–2778.
23. Pui CH. Childhood leukemias. New York: Cambridge University Press; 2006.
24. Hann I, Viscoli C, Paesmans M, et al. A comparison of outcome from febrile neutropenic episodes in children compared with adults: Results from four EORTC studies. International Antimicrobial Therapy Cooperative Group (IATCG) of the European Organization for Research and Treatment of Cancer (EORTC). Br J Haematol 1997;99:580–588.
25. Eyrich M, Wiegering V, Lim A, et al. Immune function in children under chemotherapy for standard risk acute lymphoblastic

leukaemia—A prospective study of 20 paediatric patients. Br J Haematol 2009;147:360–370.
26. Conter V, Arico M, Valsecchi MG, et al. Intensive BFM chemotherapy for childhood ALL: Interim analysis of the AIEOP-ALL 91 study. Associazione Italiana Ematologia Oncologia Pediatrica. Haematologica 1998;83:791–799.
27. Phillips B, Selwood K, Lane SM, et al. Variation in policies for the management of febrile neutropenia in United Kingdom Children's Cancer Study Group centres. Arch Dis Child 2007;92:495–498.
28. Hilf M, Yu VL, Sharp J, et al. Antibiotic therapy for Pseudomonas aeruginosa bacteremia: Outcome correlations in a prospective study of 200 patients. Am J Med 1989;87:540–546.
29. Micek ST, Lloyd AE, Ritchie DJ, et al. Pseudomonas aeruginosa bloodstream infection: Importance of appropriate initial antimicrobial treatment. Antimicrob Agents Chemother 2005;49:1306–1311.
30. Goorin AM, Hershey BJ, Levin MJ, et al. Use of trimethoprim-sulfamethoxazole to prevent bacterial infections in children with acute lymphoblastic leukemia. Pediatr Infect Dis 1985;4:265–269.
31. Hughes WT, Rivera GK, Schell MJ, et al. Successful intermittent chemoprophylaxis for Pneumocystis carinii pneumonitis. N Engl J Med 1987;316:1627–1632.
32. Schroder H, Agger KE, Rosthoj S, et al. Antibacterial prophylaxis with trimethoprim-sulfamethoxazole during induction treatment for acute lymphoblastic leukemia. Dan Med Bull 2001;48:275–277.
33. van de Wetering MD, de Witte MA, Kremer LC, et al. Efficacy of oral prophylactic antibiotics in neutropenic afebrile oncology patients: A systematic review of randomised controlled trials. Eur J Cancer 2005;41:1372–1382.
34. Takahashi N, Kameoka Y, Yamanaka Y, et al. Itraconazole oral solution enhanced vincristine neurotoxicity in five patients with malignant lymphoma. Intern Med 2008;47:651–653.
35. Venkatakrishnan K, von Moltke LL, Greenblatt DJ. Effects of the antifungal agents on oxidative drug metabolism: Clinical relevance. Clin Pharmacokinet 2000;38:111–180.
36. Leventakos K, Lewis RE, Kontoyiannis DP. Fungal infections in leukemia patients: How do we prevent and treat them? Clin Infect Dis 2010;50:405–415.
37. Marr KA, Carter RA, Crippa F, et al. Epidemiology and outcome of mould infections in hematopoietic stem cell transplant recipients. Clin Infect Dis 2002;34:909–917.
38. Lehrnbecher T, Bernig T, Hanisch M, et al. Common genetic variants in the interleukin-6 and chitotriosidase genes are associated with the risk for serious infection in children undergoing therapy for acute myeloid leukemia. Leukemia 2005;19:1745–1750.
39. Middleton PG, Cullup H, Dickinson AM, et al. Vitamin D receptor gene polymorphism associates with graft-versus-host disease and survival in HLA-matched sibling allogeneic bone marrow transplantation. Bone Marrow Transplant 2002;30:223–228.
40. Shamim Z, Ryder LP, Heilmann C, et al. Genetic polymorphisms in the genes encoding human interleukin-7 receptor-alpha: Prognostic significance in allogeneic stem cell transplantation. Bone Marrow Transplant 2006;37:485–491.
41. Kearney SL, Dahlberg SE, Levy DE, et al. Clinical course and outcome in children with acute lymphoblastic leukemia and asparaginase-associated pancreatitis. Pediatr Blood Cancer 2009;53:162–167.
42. Mitchell LG, Andrew M, Hanna K, et al. A prospective cohort study determining the prevalence of thrombotic events in children with acute lymphoblastic leukemia and a central venous line who are treated with L-asparaginase: Results of the Prophylactic Antithrombin Replacement in Kids with Acute Lymphoblastic Leukemia Treated with Asparaginase (PARKAA) Study. Cancer 2003;97:508–516.
43. Biondi A, Baruchel A, Hunger S, et al. The Eleventh International Childhood Acute Lymphoblastic Leukemia Workshop Report: Ponte di Legno, Italy, 6–7 May 2009. Leukemia 2009.

Paper II

Host genome variations and risk of infections during induction treatment of childhood acute lymphoblastic leukaemia*

Bendik Lund[1,2†], Agata Wesolowska-Andersen[3†], Birgitte Lausen[4], Louise Borst[4], Kirsten Kørup Rasmussen[4], Klaus Müller[4], Helge Klungland[2], Ramneek Gupta[3] and Kjeld Schmiegelow[4,5]

[1]Department of Paediatrics, St. Olavs Hospital, Trondheim, Norway; [2]Department of Laboratory Medicine, Children's and Women's Health, Faculty of Medicine, Norwegian University of Science and Technology, Trondheim, Norway; [3]Center for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark; [4]Department of Paediatrics and Adolescent Medicine, The University Hospital, Rigshospitalet, Copenhagen, Denmark; [5]The Institute of Gynaecology, Obstetrics and Paediatrics, The Faculty of Health Sciences, The University of Copenhagen, Copenhagen, Denmark

† Joint 1st authorship

Correspondence to: Kjeld Schmiegelow, Department of Paediatrics, The Juliane Marie Centre, The University Hospital Rigshospitalet, Blegdamsvej 9, DK-2100 Copenhagen, Denmark. Telephone: +45 35451357, Fax: +45 35454524, e-mail: Kjeld.Schmiegelow@regionh.dk

Running short title: Genomics and infections in leukaemia

Abstract word count: 204 **Text word count:** 4016

Figures: 4 **Tables:** 2 **References:** 47

Figure legends: After references

Supplemental online material: 5 figures and 5 tables

* **Financial support:** This study received financial support from the Danish Childhood Cancer Foundation; the Danish Medical Research Council for Health and Disease; The Danish Cancer Society; The Novo Nordic Foundation; The Liaison Committee Central Norway Regional Health Authority, Norwegian University of Science and Technology (NTNU); The Norwegian Childhood Cancer Organisation, Trøndelag County, Norway (grant to Bendik Lund).

ABSTRACT

Objectives

To investigate association of host genomic variation and risk of infections during treatment of childhood acute lymphoblastic leukaemia (ALL).

Methods

We explored association of 34 000 single nucleotide polymorphisms (SNPs) related primarily to pharmacogenomics and immune function to risk of infections among 69 ALL patients on induction therapy.

Results

Forty-eight (70%) patients experienced infectious events including 23 with positive blood cultures. Infectious events and positive blood cultures associated significantly with 24 and 21 SNPs, respectively ($P < 0.01$). Classification and regression tree analysis demonstrated rs11033797 (*OR51F1*), rs2835265 (*CBR1*), rs28627172 (*POLDIP3*) and rs1129844 (*CCL11*) to be predictive of outcome. Among 61 patients for whom readouts were available for all four SNPs, 40 of 41 patients with the worst SNP profile experienced at least one infectious event compared to five of the remaining 20 patients (Hazard ratio 9.0, 95% CI 3.4-23.5, which was unchanged after adjustments for neutrophil counts). Pathway analysis identified variations in 'G-protein-coupled receptor (GPCR) downstream signalling', 'Bile acid and bile salt metabolism' and 'Class I MHC mediated antigen processing & presentation' to be highly predictive of infections.

Conclusions: Our data indicate that host genomic profiling may predict the risk of infections during induction therapy. This may facilitate development of individualised supportive care.

Keywords: childhood leukaemia, SNPs, infection, immunogenetics, pharmacogenetics

INTRODUCTION

Infections remain a significant challenge in the treatment of childhood acute lymphoblastic leukaemia (ALL) (1-6) with known risk factors being use of central venous catheters (7), mucositis (8), neutropenia (9), and treatment intensity (10). Inherited genetic variation may influence the immune-inflammatory response in patients with compromised immune function such as children undergoing ALL treatment. A few studies testing single SNPs (single nucleotide polymorphisms) have shown associations between genetic variation and severity of infections during childhood leukaemia treatment both within immunogenetics (11-13) and pharmacogenetics (14). However, in children who were otherwise healthy prior to the diagnosis of ALL, severe infections are unlikely to be determined by single SNPs, since the complex factors precipitating these infections involve interplay of multiple genes and their signalling molecules (15, 16). In addition, significant single SNP associations generally have only mild effects on outcome (17). Genome-wide association studies (GWAS) provide the opportunity to link new genes to well-defined outcomes such as response to treatment, toxicities or other events. A limitation of GWAS is, however, that the majority of SNPs included on the commercially available microarray platforms are located outside protein-coding regions. Consequently, many significant SNPs in such studies are difficult to map to a specific gene and the functions of the significant SNPs are thus unclear (18, 19). A hypothesis-driven extended candidate gene approach including pathway-analysis and protein-protein interactions can provide more insights into the complex interplay between genes in complex diseases. We have recently developed a cost-effective next-generation sequencing capture assay for SNP analysis allowing pooling and simultaneously genotyping of several samples for a large number of variants (20). Aiming at exploring host genomic patterns of possible influence on infections during induction therapy, we expanded our earlier study on mannose binding lectin gene (*MBL*) polymorphisms and infections (2) to include

approximately 34 000 target SNPs in candidate genes of possible relevance for childhood ALL treatment efficacy, immune-inflammatory responses and toxicity.

MATERIAL AND METHODS

Study population

The patient population has previously been described in detail (2). Briefly, the included patients were i) diagnosed and treated according to the NOPHO ALL-92 protocol (21, 22) for non-B ALL at The University Hospital Rigshospitalet, Copenhagen, Denmark, from January 1st, 1992, to December 31st, 2000, and ii) between 1.0 -14.9 years of age at diagnosis. A total of 137 patients fulfilled these criteria and were included in the earlier study of *MBL* polymorphisms in relation to infections during the first 50 days of antileukaemic therapy (2). Of these, blood samples for extended SNP profiling were available for 69 patients who then constituted the study cohort. There were no significant differences between included and excluded patients with regard to age, risk group, immunophenotype (Table 1), proportion of patients with infectious events, and time to first infectious event from start of treatment (data not shown). With regard to gender, there was a significantly higher proportion of boys among the included patients (74%) compared to the excluded patients (53%, $P = 0.01$, Chi-Square test). Patients were treated according to the following risk criteria: a) low-risk criteria included white blood cell count (WBC) at diagnosis < $50x10^9$/L and no high-risk criteria; b) high-risk criteria included at least one of the following: i) WBC $\geq$ 50 $x10^9$/L, ii) T-cell disease, iii) a mediastinal mass, iv) t(9;22) and/or t(4;11), v) presence of CNS-disease and/or testicular involvement and/or lymphomatous leukaemia and vi) a M3 bone marrow day 15 and/or M2 bone marrow day 29.

The four-week induction therapy and three-week post-induction period consisted of prednisone 60 mg/m^2/d days 1-36, then tapered, weekly i.v. vincristine 2.0 mg/m^2 (six doses, maximum 2.0 mg per dose), i.v. doxorubicin 40 mg/m^2 (days 1, 22 and 36; with the addition of one dose on day 8 for high-risk patients), Erwinia L-Asparaginase 30 000 IE/m^2 daily days 36-45, and age-adjusted intrathecal methotrexate (four doses) (21). All patients received prophylactic trimethoprim-sulfamethoxazol, 20-25 mg/kg/d at least 2 days a week during induction treatment. Follow-up of infectious complications was done at Rigshospitalet only. Thus, no infectious episodes were missed.

Fever was defined as a single temperature measurement above 38.5°C. Neutropenia was defined as an absolute neutrophil count (ANC) of less than $0.5x10^9$/L. Since ANC was measured irregularly throughout the study period leaving many days without ANC measures, a weighing procedure was performed: the ANC measured on a specific day was repeated the next day and the day after if there was no count(s) measured the following day(s), but repeated only for a maximum of two days (giving at most three entries per single ANC measurement). Based on these extended numbers of ANC, the mean was calculated giving a weighted mean of ANC (wmANC) per patient. Similar procedures have earlier been used (23). In this retrospective analysis, an infectious event was defined as the combination of fever or other signs of infection and initiation of antimicrobial therapy after the initiation of antileukaemic therapy. Positive blood culture was defined as the isolation of a microorganism from blood. Data for each new infectious event were collected for the first 50 days of treatment.

Clinical endpoints and grouping of patients

Clinical parameters were collected from patient files and included all blood counts, days of neutropenia, febrile episodes, infectious events and microbiological data. For SNP associations, patients were grouped twice. The first stratification included cases who had at least one infectious event (n=48) during the seven weeks of antileukaemic therapy, while the remaining 21 infection-free patients served as controls. The second stratification included cases from the first grouping who in addition had at least one positive blood culture (n=23) with the remaining 46 patients serving as controls. Since ALL patients with fever were considered a heterogeneous group spanning from cases without true infections to severely ill cases with sepsis we used "culture positive" as a criteria for those with a potentially more severe febrile episode.

Candidate genes and SNP selection

Polymorphisms included in our multiple SNP analysis platform belonged to 2,350 known genes with possible relevance for childhood leukaemia treatment efficacy and toxicity (20) covering the following areas/domains: i) pharmacogenetics, ii) immunogenetics, iii) apoptosis, iv) organ specific toxicities (including thrombosis/bleeding, vincristine neuropathy, pancreatitis), v) cell cycle control genes and vi) DNA repair and mitosis. Targeted SNPs within these candidate genes were selected based on the influence on their transcript (according to Ensembl (24) annotations): i) non-synonymous coding, ii) frame-shift coding, iii) regulatory region, iv) stop lost, v) stop gained, vi) splice site, vii) within non-coding gene, and viii) within mature micro-RNA. This resulted in selection of approximately 34 000 targeted SNPs assayed in this study.

Library preparation

Blood samples were obtained during morphological remission and DNA was extracted and purified by sodium chloride and ethanol precipitation. Library preparation was performed according to the SureSelect Target Enrichment System protocol version 1.2 April 2009 (Agilent Technologies, Santa Clara, CA, USA) as described earlier (20). Briefly, three µg genomic DNA was sheared by Covaris S2 System (Covaris Inc., Woburn, MA, USA) followed by DNA purification and end-repair. For multiplexing, patient-specific barcodes of four bases were ligated to the DNA fragments. The library samples were then pooled in groups of three to eight samples and hybridized to custom designed baits (Sure Select Oligo Capture Library, Agilent Technology) with two baits targeting each SNP. After PCR amplification the library were sequenced using Illumina HiSeq 2000 (Illumina Int., San Diego, CA, USA).

Sequence analysis

The high quality reads obtained from sequencing were mapped to the reference human genome build 37 (GRCh37) using Burrow-Wheelers Alignment (BWA) algorithm (25). The alignment was refined by means of quality score recalibration and around indel realignment using Genome Analysis ToolKit (26). SNP calling was performed with SAMtools (27) using default settings. The variants were filtered using the vcfutils.pl script from SAMtools package and only genotypes at a minimum of 10x sequencing depth were included in the analysis. The high quality of SNP calling with our multiplexing technique has earlier been validated with various PCR methods (20). Variant annotation was done with Ensembl Variant Effect Predictor script (28). The data was converted into .ped and .map file formats readable by Plink (29) using vcftools (30).

Single SNP association

The single SNPs associations to risk of infection and of having a positive culture were performed by Fisher's exact test implemented in PLINK. Only SNPs with observed minor allele frequency (MAF) above 5% and at least 50% of non-missing genotypes above 10x sequencing depth were included in the analysis. A non-strict cut-off of 50% of non-missing genotypes was chosen due to novelty of this genotyping technique which often is complicated by presence of genomic regions hard to target with capture baits or difficult to sequence. The obtained *P*-values were adjusted for multiple testing with up to one million adaptive permutations (31). Finally, the adjusted *P*-values were plotted on quantile-quantile (QQ) plots (Sup. Figs 1a and 1b) using publicly available R script (http://gettinggeneticsdone.blogspot.com/) and on Manhattan plot (Fig 1) using Circos software version 0.52 (32).

Pathway analysis

All potentially functional (non-synonymous coding, frame-shift coding, stop codon and splice site) SNPs genotyped in this study with MAF > 0.01 residing in the pathways genes were retrieved for Reactome pathways (33) excluding the top two pathway levels. Number of SNPs per pathway ranged from 1 to 82 SNPs and each SNP was encoded by three values between 0 and 1 corresponding to likelihood of each genotype calculated from VCF file produced by SAMtools (34). Missing genotype calls were encoded as observed population frequencies for the three genotypes. Associations to infectious event and positive blood culture were performed by training artificial neural networks on subsets of SNPs from each pathway with 3-fold cross validation. For each pathway all combinations of up to three SNPs were assessed and the tested combinations of SNPs were ranked by Matthew's correlation coefficient (MCC) calculated as: $MCC= (TP*TN-FP*FN)/\sqrt{((TP+FP)(TP+FN)(TN+FP)(TN+FN))}$, where TP is the number of true positives, TN the number of true negatives, FP the number of

false positives and FN the number of false negatives. For each resulting best combination of SNPs for each pathway, the neural network parameters were optimized by testing multiple settings of hidden neurons and training cycles. Pathways were then ranked by MCC of the best combination of SNPs for each pathway. For all best combinations the area under receiver-operator curve (AUC) was calculated with 95% confidence intervals with 'pROC' R package.

Classification and regression tree (CART)

CART analyses were performed using rpart R package applying 3-fold cross-validation (35) using the genotypes of the SNPs associated to risk of infection and positive culture with *P*-values below 0.01 together with the patients' age, sex, risk group, immunophenotype and WBC at diagnosis. The same analysis was repeated including also pathway-based predictions achieved from top ten artificial neural network classifiers from pathway analysis.

Additional statistics

In addition to the above mentioned bioinformatic tools, Chi-Square or Fisher's Exact Test were used for univariable analysis. For "time-to-event" analysis, Kaplan-Meier analysis and log-rank test were performed for univariable comparisons, and proportional Hazard ratios were calculated for multiple regression analysis. Software package used was IBM SPSS statistics version 20.

Ethics

The study was approved by the Committee for Research Ethics in the Danish Region H and performed in accordance with the Declaration of Helsinki.

RESULTS

Of the 69 patients included in the SNP cohort, 48 (70%) patients experienced at least one infectious event (of a total of 83 events) during the seven weeks induction period with a median time to infection of 2.5 days (50% range: 1.0-12.8 days) (Table 1). Of these, 23 (33% of all) patients had at least one positive blood culture during an infectious event including a total of 40 microbiological isolates (Sup.Table 1). Of the 83 infectious events, a positive blood culture was found in 29 events (35%). Patients with infectious events were significantly younger than patients with no infectious events. Neither sex, nor immunophenotype or risk group were related to risk of infectious event (Table 1). A total of 24 and 21 SNPs were found to be significantly associated with infectious event and positive culture, respectively ($P < 0.01$ for all associations) (Sup. Tables 2 and 3). The QQ plots for the single SNP association analysis followed the null distribution, but showed a few SNPs above the expected distributions, suggesting true biological significance (Sup. Figs. 1a and 1b). Manhattan plots for both analyses indicate that several loci were associated with both risk of having an infection and positive culture (Fig. 1). The most significant SNP associated with infectious event was a synonymous coding SNP (rs11033797) in *OR51F1*, while the top SNP associated with positive blood culture was a non-synonymous coding SNP (rs12632456) in the *FLNB* gene. Pathway analysis identified 'G-protein-coupled receptor (GPCR) downstream signalling' (MCC = 0.72, AUC = 0.88), 'Bile acid and bile salt metabolism' (MCC = 0.71, AUC = 0.89), and 'Class I MHC mediated antigen processing & presentation' (MCC = 0.68, AUC = 0.83) among the pathways most predictive of infectious event. 'Interferon Signalling' (MCC = 0.70, AUC = 0.85), 'Rho GTPase cycle' (MCC = 0.62, AUC = 0.86) and 'G alpha (i) signalling events' (MCC = 0.60, AUC = 0.85) pathways were the most predictive for having positive blood culture. The top ten pathways together with selected combinations of SNPs for risk of infectious event are listed in Sup. Table 4 and for positive cultures in Sup. Table 5.

CART analysis demonstrated rs11033797 (SNP1; *OR51F1*), rs2835265 (SNP2; *CBR1*), rs28627172 (SNP3; *POLDIP3*) and rs1129844 (SNP4; *CCL11*) to be highly predictive of infectious event (Fig. 2a). Patients were classified as "SNP profile-positive" (P+) according to the following SNP-pattern: either i) SNP1 (GA, GG) or ii) SNP1 (AA) and SNP2 (TC), or iii) SNP1 (AA) and SNP2 (CC) and SNP3 (GG) and SNP4 (GG). The remaining patients were classified as "SNP profile negative" (P-). Data for all four SNPs in the profile were available for 61 out of 69 patients. Of these 61 patients, 45 experienced at least one infectious event. Of the P+ patients, 40 (97.6%) out of 41 experienced an event compared to only 5 (25%) out of the 20 P- patients (Fig. 2b). In time-to-event analysis, the risk of having an infectious event for P+ patients compared to P- patients was 9.0 (95% CI: 3.4-23.5, Hazard Ratio, Cox regression, Table 2). Adjusting for wmANC and age did not change the risk for P+ patients significantly (Table 2). Age demonstrated to be an independent risk factor for having an infectious event with increased risk in the lowest age group (Table 2). There was no significant interaction between age and SNP-profile; hence, the effect of SNP-profile as a risk factor was not influenced by age. For the "culture positive" patients (n=23), CART analysis was less predictive and rs12632456 (*FLNB*) and rs1171218 (*TOPBP1*) identified only 10 of 23 (43%) culture positive patients with 77% accuracy (Sup. Fig. 2).

CART analysis including pathway profiles demonstrated a combination of variations in 'GPCR downstream signalling', 'Bile acid and salt metabolism' and 'Class I MHC antigen processing & presentation" to be highly predictive of infectious event, while combinations of variations within 'Interferon signalling' and 'Platelet Aggregation (Plug Formation)' pathways were predictive of a positive blood culture (Sup. Figs. 3a and 3b).

DISCUSSION

The cure rates of childhood ALL are approaching 85-90%. Of the 10-15% of patients that die, as many as 25% do so from treatment related toxicity rather than from active cancer (1, 36). Since relapse rates are decreasing because of better risk adapted, but also more intensive therapy, focus on toxic events has become increasingly important. Among patients dying from toxicity, infections are the most common cause and the majority of these occur during the first months of treatment (1, 4). It remains unknown why some patients develop severe, even life-threatening infections during treatment, while others, within the same sex, age and risk group only experience mild infections. Many factors influence the risk of severe infections during ALL treatment (Fig. 3). However, the role of inherited genetic factors has to date not been investigated extensively.

The present study strongly indicates that common host genomic variants influencing immune function, drug disposition and other intracellular signalling mechanisms may play a critical role. However, infectious events constitute a clinically heterogeneous group spanning from mild clinical episodes of fever of unknown origin to severe infectious episodes requiring intensive care. Accordingly, the retrospective nature of this study and the definition of an infectious event (fever or other signs of infection and start of antibiotics) may have both under- and overestimated the impact of host genomics.

Because of the complex nature of infections, a multiple gene and SNP profiling approach is more likely to yield biologically relevant results compared to single SNP analysis. Importantly, our findings of 24 and 21 SNPs significantly associated with the risk of an infectious event or a positive culture, respectively, emphasize that application of such a candidate gene approach can yield potentially relevant results, not least when the SNP data

are integrated into pathway explorations. Still, the results of this study need to be validated in larger independent cohorts.

Since our SNP risk profile is based on a mathematical tool as CART analysis, the combination of the most predictive SNPs does not necessarily reflect true biological coherence. In addition, a single predictive SNP might not be causative because it can be in linkage disequilibrium (i.e. a tag SNP) with the causative SNP which may not has been included in our SNP-panel. The four predictive SNPs in our CART analysis link four genes to the risk of having an infectious event. The *OR51F1*-gene belongs to the superfamily of olfactory receptors (OR) which constitute the largest class of human G-protein coupled receptors (GPCRs), and currently about 900 human OR genes are known (37). Theoretically, there is some functional analogy between germinal centres B-lymphocytes and differentiating olfactory sensory neurons (38), but any relationship with leukaemia or infections are unknown. The SNP was included in our panel because it was located in an intron of another gene (*MMP26*) involved in tumour invasion and metastasis. One can speculate whether there is a hitherto not detected link between *OR51F1* and risk of infections, or that the SNP only is a tag-SNP with no actual biological relation to infections. Alternatively, it could also be a false positive finding potentially caused by mutational heterogeneity frequently seen particularly for olfactory receptors (39). The *CBR1*-gene encodes a carbonyl reductase which is a NADPH-dependent oxidoreductase with specificity for carbonyl compounds. There is some evidence that *CBR1* influences anthracycline metabolism (40) which is part of the induction treatment. Patients with possibly reduced metabolism of doxorubicin might have increased toxicity resulting in prolonged neutropenia, and thereby increased risk of infections. The *POLDIP3*-gene encodes an enzyme (polymerase delta-interacting protein 3) which interacts with DNA polymerase regulating cell growth. It has been associated with sleep

patterns in childhood ALL patients during dexamethasone treatment (41) and autoimmune disorders (42). An association with infections in immunocompromised patients has not earlier been reported. The *CCL11*-gene encodes a chemokine, eotaxin-1, which is a member of the CC subfamily and displays chemotactic activity to eosinophil's. Its known clinical function relates to asthma, allergic diseases and parasite infections (43), but as an inflammatory signalling molecule, a role in host defence towards infections in our patients cannot be excluded. For the other end point in our study, "culture positive" patients, the most predictive SNP was located in the *FLNB*-gene which encodes Filamin B, a cytoplasmic protein binding actin and connecting plasma membrane with the intracellular cytoskelleton, thereby changing the cell shape. It is expressed in vessels and endothelial cells and takes part in the process of angiogenesis (44). It resembles Filamin A which has been associated with increased vascular permeability (45) known to increase risk of severe sepsis. Whether this also applies to Filamin B remains unclear.

The earlier mentioned SNP/gene studies on childhood leukaemia and infections/neutropenia (11-14) included a total of 21 SNPs of which 15 were included in our SNP-analysis platform. For these SNPs, differences were found in minor allele frequencies distribution with regard to "infectious events", but they were not significant. Whether some of these SNPs might have been significant in a study with larger sample size is unknown. For the other end point, "culture positive", two of the SNPs in the study by Kidas *et al.* (11) were significantly associated with outcome (rs909253(+252A>G)/*LT-α* gene; rs1800629(-308G>A)/*TNF-α* gene), but in the analysis of sequencing data for this project they were filtered out due to quality check (base quality bias and strand bias in sequencing data). A more simplified replication study of this finding could probably have clarified the significance of these genes/SNPs.

Since our starting point is genes and SNPs with known biological functions related to childhood leukaemia, pathway analysis was made possible. Integrative analysis of effects mediated by multiple SNPs grouped by their function, e.g. acting in the same biological pathway, can provide robust results identifying the most important underlying biological mechanisms. The two pathways most predictive of infectious event, 'GPCR downstream signalling' and 'Bile acid and salt metabolism', are both connected to steroid drugs that constitute an important component of induction therapy (46) and have profound impact on immune function. Variants in the second pathway are also within genes involved in steroid metabolism, while GPCR signalling reflects the mode of action of those drugs (47). Finally, 'Class I MHC antigen processing & presentation' is a critical pathway in response to viral and other infections, as are the two pathways most predictive of positive blood culture 'Interferon signalling' and 'Rho GTPase cycle'.

Although the hierarchical approach of CART analysis reduces the true complexity of the biological and environmental (including instrumental) factors leading to infections, its simplicity makes the risk profile data more comprehensible and clinically applicable than complex pathway interactions. In our study, the identification of a 4-SNP risk profile predicting the occurrence of an infectious event for 40 out of 45 patients with 98% accuracy might easily be utilized in a clinical setting. Genotyping of patients at time of diagnosis could possibly identify patients at risk of infectious events and prophylactic actions might be considered including use of prophylactic antibiotics, closely monitoring of patients and/or other supportive care measures. In future studies, our model might also be used for identification of SNP risk profiles discriminating patients at risk of infectious death. For some patients the risk of infectious death might possibly exceed the risk of death after relapse,

hence, treatment intensity should be reduced or changed for such patients. However, a SNP-risk profile should be demonstrated predictive in a prospective study before any use in a clinical setting.

Our study had some limitations, including small number of patients and a risk of false positive results. Additionally, investigations of a set of candidate polymorphisms, even when conducted on a large scale, a priori exclude identification of other important but yet unknown biological mechanisms. Finally, DNA was available for only 50% of patients from the original cohort, but our SNP cohort differed only by gender compared to the excluded patients with an increased proportion of boys (Table 1). Importantly, gender was not identified as a risk factor for having an infectious event or a positive culture (Table 1), which is also in agreement with previous studies (13). In our earlier study on *MBL* polymorphisms there was also no gender interaction with impact of *MBL* genotypes (2), supporting the view that gender probably is of limited importance in this setting.

Despite the small size of the cohort, the associations were quite strong and the results indicate that host genomic profiling may predict the risk of infectious events during induction therapy in children with ALL. Furthermore, the integrative bioinformatic approach allows identification of critical pathways even though the impact of individual SNPs may be low. Such knowledge might be helpful in identifying patients at high risk of infectious complications and thus in need for more individualised supportive care and even antileukaemic treatment protocols. Furthermore, the study design used here can be applied for many other clinical endpoints, for instance other toxicities such as toxic deaths, pancreatitis, thrombosis or osteonecrosis. Such candidate SNP profiling and association studies are on-going as part of the Nordic/Baltic ALL 2008 protocol for childhood and adult ALL.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This study received financial support from The Danish Childhood Cancer Foundation; The Danish Medical Research Council for Health and Disease; The Danish Cancer Society; The Novo Nordic Foundation; The Liaison Committee Central Norway Regional Health Authority, Norwegian University of Science and Technology (NTNU); The Norwegian Childhood Cancer Organisation, Trøndelag County, Norway (grant to Bendik Lund).

AUTHORSHIP CONTRIBUTIONS

Bendik Lund and Kjeld Schmiegelow designed the study. Ramneek Gupta, Agata Wesolowska-Andersen, Louise Borst, Klaus Müller and Kjeld Schmiegelow identified the relevant genes and designed the SNP profiling approach. Bendik Lund and Agata Wesolowska-Andersen wrote the manuscript, performed data analysis and interpreted the data. Birgitte Lausen was responsible for collection of the clinical data. Louise Borst and Kirsten Kørup Rasmussen performed the laboratory work. Helge Klungland, Ramneek Gupta and Kjeld Schmiegelow supervised data analysis. All authors provided critical input to the project and manuscript and approved the final manuscript.

REFERENCES

1. Lund B, Asberg A, Heyman M, Kanerva J, Harila-Saari A, Hasle H, Soderhall S, Jonsson OG, Lydersen S, Schmiegelow K. Risk factors for treatment related mortality in childhood acute lymphoblastic leukaemia. Pediatric Blood & Cancer. 2011; **56**(4): 551-9.

2. Lausen B, Schmiegelow K, Andreassen B, Madsen HO, Garred P. Infections during induction therapy of childhood acute lymphoblastic leukemia--no association to mannose-binding lectin deficiency. Eur J Haematol. 2006; **76**(6): 481-7.

3. Afzal S, Ethier MC, Dupuis LL, Tang L, Punnett AS, Richardson SE, Allen U, Abla O, Sung L. Risk factors for infection-related outcomes during induction therapy for childhood acute lymphoblastic leukemia. The Pediatric Infectious Disease Journal. 2009; **28**(12): 1064-8.

4. Prucker C, Attarbaschi A, Peters C, Dworzak MN, Potschger U, Urban C, Fink FM, Meister B, Schmitt K, Haas OA, Gadner H, Mann G. Induction death and treatment-related mortality in first remission of children with acute lymphoblastic leukemia: a population-based analysis of the Austrian Berlin-Frankfurt-Munster study group. Leukemia. 2009; **23**(7): 1264-9.

5. Bailey LC, Reilly AF, Rheingold SR. Infections in pediatric patients with hematologic malignancies. Semin Hematol. 2009; **46**(3): 313-24.

6. Rahiala J, Perkkio M, Riikonen P. Infections occurring during the courses of anticancer chemotherapy in children with ALL: a retrospective analysis of 59 patients. Pediatr Hematol Oncol. 1998; **15**(2): 165-74.

7. Allen RC, Holdsworth MT, Johnson CA, Chavez CM, Heideman RL, Overturf G, Lemon D, Hunt WC, Winter SS. Risk determinants for catheter-associated blood stream infections in children and young adults with cancer. Pediatric Blood & Cancer. 2008; **51**(1): 53-8.

8. Elting LS, Cooksley C, Chambers M, Cantor SB, Manzullo E, Rubenstein EB. The burdens of cancer therapy. Clinical and economic outcomes of chemotherapy-induced mucositis. Cancer. 2003; **98**(7): 1531-9.

9. Castagnola E, Fontana V, Caviglia I, Caruso S, Faraci M, Fioredda F, Garre ML, Moroni C, Conte M, Losurdo G, Scuderi F, Bandettini R, Toma P, Viscoli C, Haupt R. A prospective study on the epidemiology of febrile episodes during chemotherapy-induced neutropenia in children with cancer or after hemopoietic stem cell transplantation. Clin Infect Dis. 2007; **45**(10): 1296-304.

10. Sung L, Gamis A, Alonzo TA, Buxton A, Britton K, Deswarte-Wallace J, Woods WG. Infections and association with different intensity of chemotherapy in children with acute myeloid leukemia. Cancer. 2009; **115**(5): 1100-8.

11. Kidas E, Moricke A, Beier R, Welte K, Schrappe M, Stanulla M, Grigull L. Genetic polymorphisms of the lymphotoxin alpha gene are associated with increased risk for lethal infections during induction therapy for childhood acute leukemia: a case-control study. Int J Hematol. 2009; **89**(5): 584-91.

12. Miedema KG, te Poele EM, Tissing WJ, Postma DS, Koppelman GH, de Pagter AP, Kamps WA, Alizadeh BZ, Boezen HM, de Bont ES. Association of polymorphisms in the TLR4 gene with the risk of developing neutropenia in children with leukemia. Leukemia. 2011; **25**(6): 995-1000.

13. Lehrnbecher T, Bernig T, Hanisch M, Koehl U, Behl M, Reinhardt D, Creutzig U, Klingebiel T, Chanock SJ, Schwabe D. Common genetic variants in the interleukin-6 and chitotriosidase genes are associated with the risk for serious infection in children undergoing therapy for acute myeloid leukemia. Leukemia. 2005; **19**(10): 1745-50.

14. Erdelyi DJ, Kamory E, Zalka A, Semsei AF, Csokay B, Andrikovics H, Tordai A, Borgulya G, Magyarosy E, Galantai I, Fekete G, Falus A, Szalai C, Kovacs GT. The role of

ABC-transporter gene polymorphisms in chemotherapy induced immunosuppression, a retrospective study in childhood acute lymphoblastic leukaemia. Cell Immunol. 2006; **244**(2): 121-4.

15. Huttunen R, Aittoniemi J. New concepts in the pathogenesis, diagnosis and treatment of bacteremia and sepsis. J Infect. 2011; **63**(6): 407-19.

16. Marshall JC, Vincent JL, Fink MP, Cook DJ, Rubenfeld G, Foster D, Fisher CJ, Jr., Faist E, Reinhart K. Measures, markers, and mediators: toward a staging system for clinical sepsis. A report of the Fifth Toronto Sepsis Roundtable, Toronto, Ontario, Canada, October 25-26, 2000. Crit Care Med. 2003; **31**(5): 1560-7.

17. Lesnick TG, Papapetropoulos S, Mash DC, Ffrench-Mullen J, Shehadeh L, de Andrade M, Henley JR, Rocca WA, Ahlskog JE, Maraganore DM. A genomic pathway approach to a complex disease: axon guidance and Parkinson disease. PLoS Genet. 2007; **3**(6): e98.

18. Yang JJ, Cheng C, Yang W, Pei D, Cao X, Fan Y, Pounds SB, Neale G, Trevino LR, French D, Campana D, Downing JR, Evans WE, Pui CH, Devidas M, Bowman WP, Camitta BM, Willman CL, Davies SM, Borowitz MJ, Carroll WL, Hunger SP, Relling MV. Genome-wide interrogation of germline genetic variation associated with treatment response in childhood acute lymphoblastic leukemia. JAMA: the journal of the American Medical Association. 2009; **301**(4): 393-403.

19. Maher B. ENCODE: The human encyclopaedia. Nature. 2012; **489**(7414): 46-8.

20. Wesolowska A, Dalgaard MD, Borst L, Gautier L, Bak M, Weinhold N, Nielsen BF, Helt LR, Audouze K, Nersting J, Tommerup N, Brunak S, Sicheritz-Ponten T, Leffers H, Schmiegelow K, Gupta R. Cost-effective multiplexing before capture allows screening of 25 000 clinically relevant SNPs in childhood acute lymphoblastic leukemia. Leukemia. 2011; **25**(6): 1001-6.

21. Gustafsson G, Schmiegelow K, Forestier E, Clausen N, Glomstein A, Jonmundsson G, Mellander L, Makipernaa A, Nygaard R, Saarinen-Pihkala UM. Improving outcome through two decades in childhood ALL in the Nordic countries: the impact of high-dose methotrexate in the reduction of CNS irradiation. Nordic Society of Pediatric Haematology and Oncology (NOPHO). Leukemia. 2000; **14**(12): 2267-75.

22. Schmiegelow K, Forestier E, Hellebostad M, Heyman M, Kristinsson J, Soderhall S, Taskinen M. Long-term results of NOPHO ALL-92 and ALL-2000 studies of childhood acute lymphoblastic leukemia. Leukemia. 2010; **24**(2): 345-54.

23. Schmiegelow K, Pulczynska MK. Maintenance chemotherapy for childhood acute lymphoblastic leukemia: should dosage be guided by white blood cell counts? Am J Pediatr Hematol Oncol. 1990; **12**(4): 462-7.

24. Hubbard T, Barker D, Birney E, Cameron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V, Down T, Durbin R, Eyras E, Gilbert J, Hammond M, Huminiecki L, Kasprzyk A, Lehvaslaiho H, Lijnzaad P, Melsopp C, Mongin E, Pettett R, Pocock M, Potter S, Rust A, Schmidt E, Searle S, Slater G, Smith J, Spooner W, Stabenau A, Stalker J, Stupka E, Ureta-Vidal A, Vastrik I, Clamp M. The Ensembl genome database project. Nucleic Acids Res. 2002; **30**(1): 38-41.

25. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010; **26**(5): 589-95.

26. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; **20**(9): 1297-303.

27. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; **25**(14): 1754-60.

28. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics. 2010; **26**(16): 2069-70.

29. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; **81**(3): 559-75.

30. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R. The variant call format and VCFtools. Bioinformatics. 2011; **27**(15): 2156-8.

31. Besag J, Clifford P. Sequential Monte Carlo *p*-values. Biometrika. 1991; **78**(2): 301-4.

32. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. Genome Res. 2009; **19**(9): 1639-45.

33. Joshi-Tope G, Gillespie M, Vastrik I, D'Eustachio P, Schmidt E, de Bono B, Jassal B, Gopinath GR, Wu GR, Matthews L, Lewis S, Birney E, Stein L. Reactome: a knowledgebase of biological pathways. Nucleic Acids Res. 2005; **33**(Database issue): D428-32.

34. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; **25**(16): 2078-9.

35. Therneau TM, Atkinson EJ. An introduction to recursive partitioning using the rpart routines. . Technical report: Mayo Clinic, Rochester. http://www.mayo.edu/hsr/techrpt/61.pdf; 1997.

36. Schmiegelow K, Al-Modhwahi I, Andersen MK, Behrendtz M, Forestier E, Hasle H, Heyman M, Kristinsson J, Nersting J, Nygaard R, Svendsen AL, Vettenranta K, Weinshilboum R. Methotrexate/6-mercaptopurine maintenance therapy influences the risk of

a second malignant neoplasm after childhood acute lymphoblastic leukemia: results from the NOPHO ALL-92 study. Blood. 2009; **113**(24): 6077-84.

37. Feldmesser E, Olender T, Khen M, Yanai I, Ophir R, Lancet D. Widespread ectopic expression of olfactory receptor genes. BMC Genomics. 2006; **7**: 121.

38. Otaki JM, Fearon DT, Yamamoto H. The proto-oncogene BCL-6 is expressed in olfactory sensory neurons. Neurosci Res. 2005; **53**(2): 189-200.

39. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA, Kiezun A, Hammerman PS, McKenna A, Drier Y, Zou L, Ramos AH, Pugh TJ, Stransky N, Helman E, Kim J, Sougnez C, Ambrogio L, Nickerson E, Shefler E, Cortes ML, Auclair D, Saksena G, Voet D, Noble M, DiCara D, Lin P, Lichtenstein L, Heiman DI, Fennell T, Imielinski M, Hernandez B, Hodis E, Baca S, Dulak AM, Lohr J, Landau DA, Wu CJ, Melendez-Zajgla J, Hidalgo-Miranda A, Koren A, McCarroll SA, Mora J, Lee RS, Crompton B, Onofrio R, Parkin M, Winckler W, Ardlie K, Gabriel SB, Roberts CW, Biegel JA, Stegmaier K, Bass AJ, Garraway LA, Meyerson M, Golub TR, Gordenin DA, Sunyaev S, Lander ES, Getz G. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013; **499**(7457): 214-8.

40. Bains OS, Karkling MJ, Grigliatti TA, Reid RE, Riggs KW. Two nonsynonymous single nucleotide polymorphisms of human carbonyl reductase 1 demonstrate reduced in vitro metabolism of daunorubicin and doxorubicin. Drug Metab lism and Disposition: The Biological Fate of Chemicals. 2009; **37**(5): 1107-14.

41. Vallance K, Liu W, Mandrell BN, Panetta JC, Gattuso JS, Hockenberry M, Zupanec S, Yang L, Yang J, Hinds PS. Mechanisms of dexamethasone-induced disturbed sleep and fatigue in paediatric patients receiving treatment for ALL. Eur J Cancer. 2010; **46**(10): 1848-55.

42. Avila J, Acosta E, Machargo MD, Arteaga MF, Gallego E, Canete H, Garcia-Perez JJ, Martin-Vasallo P. Autoantigenic nuclear proteins of a clinically atypical renal vasculitis. J Autoimmune Dis. 2008; **5**: 3.

43. Simons JE, Rothenberg ME, Lawrence RA. Eotaxin-1-regulated eosinophils have a critical role in innate immunity against experimental Brugia malayi infection. Eur J Immunol. 2005; **35**(1): 189-97.

44. Del Valle-Perez B, Martinez VG, Lacasa-Salavert C, Figueras A, Shapiro SS, Takafuta T, Casanovas O, Capella G, Ventura F, Vinals F. Filamin B plays a key role in vascular endothelial growth factor-induced endothelial cell motility through its interaction with Rac-1 and Vav-2. J Biol Chem. 2010; **285**(14): 10748-60.

45. Griffiths GS, Grundl M, Allen JS, 3rd, Matter ML. R-Ras interacts with filamin a to maintain endothelial barrier function. J Cell Physiol. 2011; **226**(9): 2287-96.

46. Gaynon PS, Carrel AL. Glucocorticosteroid therapy in childhood acute lymphoblastic leukemia. Adv Exp Med Biol. 1999; **457**: 593-605.

47. Tasker JG, Di S, Malcher-Lopes R. Minireview: rapid glucocorticoid signaling via membrane-associated receptors. Endocrinology. 2006; **147**(12): 5549-56.

FIGURE LEGENDS

Figures:

- **Figure 1.** Manhattan plot (hs1-hsY=chr.1-chr.Y) showing the association of SNPs with 1) "infectious event" (inner gray-shaded ring), and 2) "positive culture" (outer gray-shaded ring). The radius ("y-axis") represents the $-\log_{10}$ of *P*-values and each SNP is plotted based on its position in the chromosome and as a function of its *P*-value. Red dots represent SNPs with *P*-values < 0.01. Grey dots represent SNPs with *P*-values >0.01. Green tick marks represent positions of baits used for target capture.
- **Figure 2a**. CART (Classification and regression tree) diagram for «infectious event» showing the most predictive SNP-sequences (all SNPs $P < 0.01$) and the corresponding genes (SNPs termed 1-4).
- **Figure 2b.** SNP-profile and time to first infectious event. Total number of patients: **61**. Total number of events: **45**. Patients were classified as "SNP profile-positive" (P+) according to the following SNP-pattern: either i) SNP1 (GA, GG) or ii) SNP1 (AA) and SNP2 (TC), or iii) SNP1 (AA) and SNP2 (CC) and SNP3 (GG) and SNP4 (GG). The remaining patients were classified as "SNP profile-negative" (P-). The figure illustrates the risk of having an infectious event with **positive** (97.6% ± 2.4%, 40/41) vs. **negative** (25% ± 9.7%, 5/20) SNP profile ($P < 0.001$, Log Rank).
- **Figure 3.** A model illustrating possible factors influencing outcome in infectious events during ALL treatment.

Supplementary Figures:

- **Supplementary Figure 1a and 1b. QQ plots for single SNP association analysis**. The plots followed the null distribution, but showed a few SNPs above the expected distributions for both infectious events (1a) and positive culture (1b), suggesting true biological signals.
- **Supplementary Figure 2.** CART (Classification and regression tree) diagram for «culture positive » patients showing the two most predictive SNP-sequences (all SNPs $P < 0.01$) and the corresponding genes.
- **Supplementary Figure 3a.** CART analysis for risk of infectious events including pathway profiles. F=favorable profile, U=unfavorable profile.
- **Supplementary Figure 3b.** CART analysis for risk of positive blood culture including pathway profiles. F=favorable profile, U=unfavorable profile.

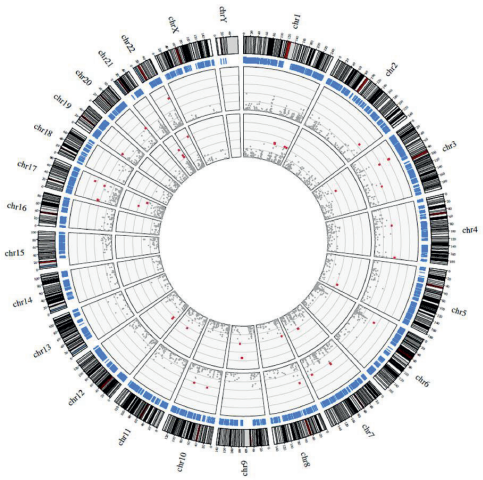


Figure 1.

Figure 2a. CART diagram: infectious event

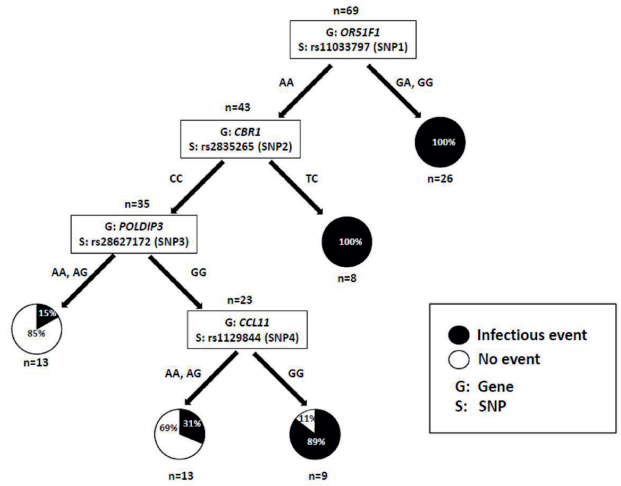


Figure 2b. SNP-profile and time to first infectious event.

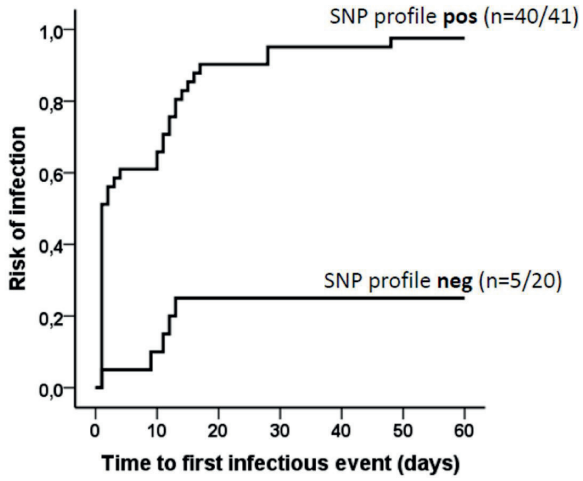


Figure 3. A model illustrating possible factors influencing outcome in infectious events during ALL treatment.

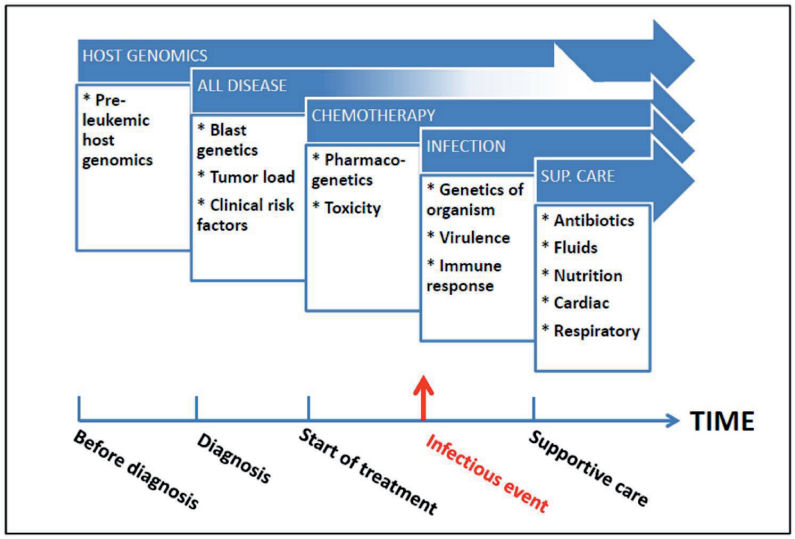


Table 1. Patient characteristics of included patients (n=69) compared with original cohort (n=137)

Patients	Original cohort (n=137) No. (%)	SNP cohort (n=69) No. (%)	p-value, included vs. excluded[a]	Inf. event, (n=48) No. (%)[b]	p-value, infectious event[b]	Pos. culture, (n=23) No. (%)[b]	p-value, positive culture[b]
Sex			**0.01***		0.78		0.56
Female	50 (37)	18 (26)		13 (27)		7 (30)	
Male	87 (63)	51 (74)		35 (73)		16 (70)	
Age			0.09		**<0.01***		0.24
1-5	86 (63)	49 (71)		41 (85)		18 (78)	
6-10	32 (23)	14 (20)		6 (13)		5 (22)	
11-14	19 (14)	6 (9)		1 (2)		0 (0)	
Risk group			0.06		0.12		0.31
Low risk	95 (69)	53 (77)		34 (71)		16 (70)	
High risk	42 (31)	16 (23)		14 (29)		7 (30)	
Immunoph.			0.18		0.71		0.47
Non-B cell	113 (83)	60 (87)		41 (85)		19 (83)	
T-cell	24 (17)	9 (13)		7 (15)		4 (17)	

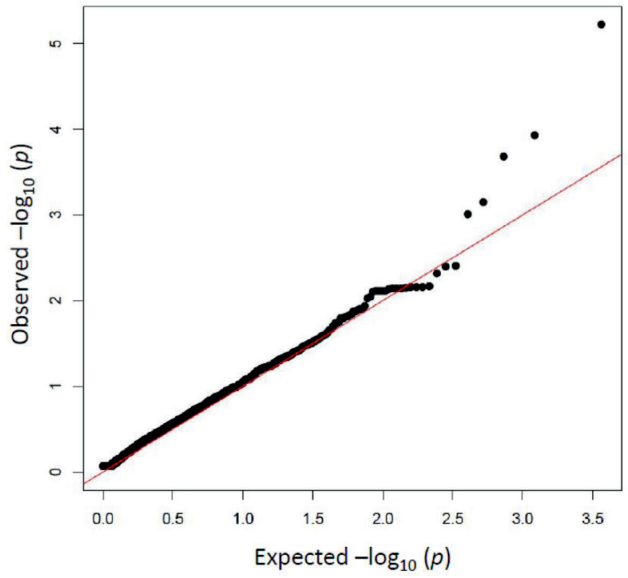
[a] SNP cohort (=included patients) vs. excluded patients
[b] Comparison of patient characteristics between events-patients (infectious event/positive culture) and no-events patients, SNP cohort only (Pearson Chi-Square/ Fisher's exact tests).
* significant

Table 2. Multiple regression analysis showing risk factors for having an infectious event.

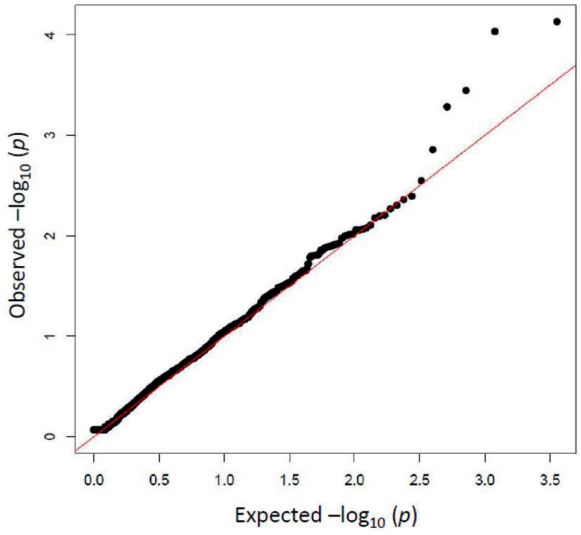
Risk factor	Events n=45	All patients n=61	HR (95% CI) Simple regr.	Adjusted HR (95% CI) Multiple regr.
SNP profile				
Positive	40	41	**9.0 (3.4-23.5)***	**7.5 (2.8-20.1)***
Negative	5	20	1.0	1.0
wmANC				
<median	25	32	1.2 (0.7-2.2)	0.8 (0.5-1.5)
>median	20	29	1.0	1.0
Age				
1-5	38	43	1.0	1.0
6-10	6	13	**0.3 (0.1-0.7)***	**0.3 (0.1-0.8)***
11-14	1	5	**0.1 (0.0-0.8)***	0.2 (0.0-1.6)

HR: Hazard ratio
wmANC: weighted mean of absolute neutrophile count (ANC)
Only patients with sufficient SNP profile sequenced was included
*significant

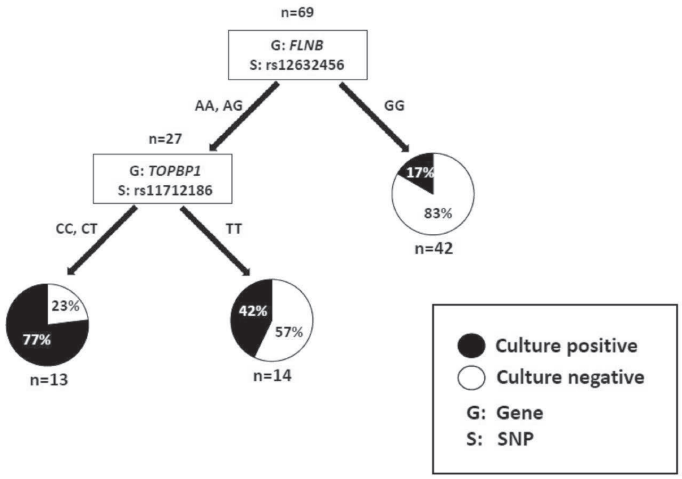
Supplementary Figure 1a. QQ-plot for infectious event.



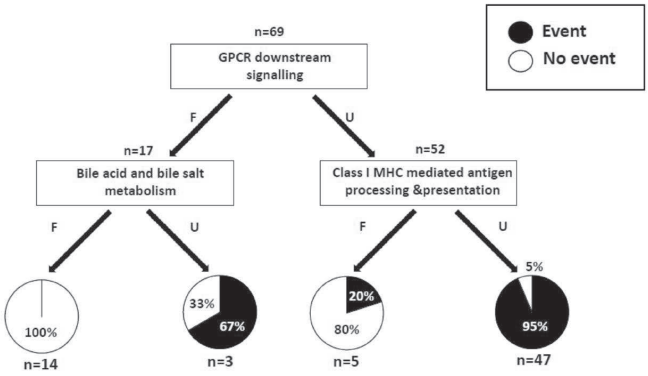
Supplementary Figure 1b. QQ-plot for positive culture .



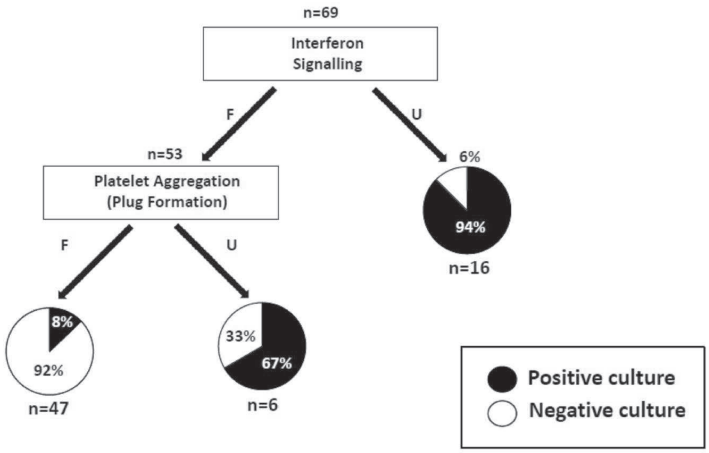
Supplementary Figure 2.
CART diagram: culture positive



Supplementary Figure 3a.



Supplementary Figure 3b.



Supplementary Table 1. Blood cultures/isolates[a] identified in culture positive patients.

Microbiological agens	No of isolates (n=40)
Streptococcus species[b]	8
Staphylococcus species [c]	14
Pseudomonas species[d]	5
Candida albicans	3
Escherichia coli	6
Enterobacter species	1
Actineobacter	1
Campylobacter	1
Pantoea species	1

[a] A total of 40 isolates were found in the 23 patients with at least 1 positive culture .
[b] Isolated Streptococcus species included: non-haemolytic, faecalis, oralis, salivarius, mitis and sanguis
[c] Isolated Staphylococcus species included: epidermidis, coagulase negative, hominis and aureus
[d] Isolated Pseudomonas species included: aeruginosa, stutzeri and maltofilia

Supplementary Table 2. SNPs associated with risk of infectious event with permutation corrected *P*-values < 0.01

rsID	Perm p-val	Genotypes – no event	Genotypes - event	SNP consequence	HGNC
rs11033797	6.00E-06	GG:0;GA:0;AA:20;00:1	GG:1;GA:25;AA:22;00:0	SYNONYMOUS_CODING	*OR51F1*
rs28627172	0.0001	AA:2;AG:8;GG:8;00:3	AA:0;AG:5;GG:40;00:3	NON_SYNONYMOUS_CODING	*POLDIP3*
rs35158358	0.0002	-/-:0;-/+C:9;+C/+C:10;00:2	+C/+C:6;+C/-:26;-/-:12/00:4	WITHIN_NON_CODING_GENE	*GDA*
rs73197348	0.0007	CC:0;CT:7;TT:13;00:1	CC:0;CT:2;TT:46;00:0	REGULATORY_REGION	*RUNX1*
rs1129844, CM072923	0.001	AA:3;AG:7;GG:6;00:5	AA:0;AG:10;GG:34;00:4	NON_SYNONYMOUS_CODING	*CCL11*
rs2835265	0.004	TT:0;TC:0;CC:20;00:1	TT:0;TC:14;CC:34;00:0	NON_SYNONYMOUS_CODING	*CBR1*
rs11409972	0.004	+T/+T:1;+T/-:7; -/-:10;00:3	+T/+T:9;+T/-:24;-/-:12;00:3	REGULATORY_REGION	*MAGI1*
rs67155431	0.005	-/-:1;-/+A:4;+A/+A:7;00:9	-/-:0;-/+A:4;+A/+A:38/00:6	SPLICE_SITE	*SLCO1A2*
rs10653020	0.007	+CATC/+CATC:0;+CATC/-:1;-/-:17;00:3	+CATC/+CATC:0;+CATC/-:17;-/-:27;00:4	WITHIN_NON_CODING_GENE	*ACE*
rs1130435	0.007	TT:1;TC:5;CC:10;00:5	TT:8;TC:24;CC:10;00:6	NON_SYNONYMOUS_CODING	*FABP6*
rs76959009	0.007	CC:0;CT:2;TT:18;00:1	CC:0;CT:20;TT:26;00:2	WITHIN_NON_CODING_GENE	*C1orf112*
rs1136410, CM042761	0.007	GG:0;GA:8;AA:3;00:10	GG:0;GA:10;AA:27;00:11	NON_SYNONYMOUS_CODING	*PARP1*
rs35710857	0.007	TT:0;TC:0;CC:21;00:0	TT:0;TC:13;CC:35;00:0	SYNONYMOUS_CODING	*CBR1*
rs2292572	0.007	TT:0;TG:2;GG:18;00:1	TT:5;TG:15;GG:28;00:0	REGULATORY_REGION	*GAB2*
rs2275287	0.007	TT:1;TC:4;CC:9;00:7	TT:8;TC:27;CC:9;00:4	SPLICE_SITE	*RYR2*
rs10841795	0.007	GG:0;GA:5;AA:10;00:6	GG:0;GA:3;AA:43;00:2	NON_SYNONYMOUS_CODING	*SLCO1A2*
rs12666401	0.007	AA:0;AG:0;GG:14;00:7	AA:0;AG:14;GG:27;00:7	REGULATORY_REGION	*SHFM1*
rs5367	0.008	GG:0;GA:2;AA:17;00:2	GG:0;GA:20;AA:25;00:3	SPLICE_SITE	*SELE*
rs12133666	0.008	TT:0;TA:2;AA:19;00:0	TT:0;TA:20;AA:28;00:0	WITHIN_NON_CODING_GENE	*SELE*
rs3917441	0.008	CC:0;CG:2;GG:18;00:1	CC:0;CG:20;GG:28;00:0	REGULATORY_REGION	*SELE*
rs77115118	0.008	TT:0;TA:2;AA:19;00:0	TT:0;TA:20;AA:28;00:0	WITHIN_NON_CODING_GENE	*SELE*
rs62471402	0.008	GG:0;GC:0;CC:15;00:6	GG:0;GC:14;CC:33;00:1	WITHIN_NON_CODING_GENE	*SHFM1*
rs4877837	0.009	AA:0;AG:6;GG:3;00:12	AA:0;AG:7;GG:26;00:15	REGULATORY_REGION	*SLC28A3*
rs2306825	0.009	AA:0;AG:1;GG:16;00:4	AA:0;AG:17;GG:28;00:3	SYNONYMOUS_CODING, SPLICE_SITE	*PSD3*

Supplementary Table 3. SNPs associated with positive blood culture with permutation corrected *P*-values < 0.01.

rsID	Perm p-val	Genotypes – positive culture	Genotypes – no positive culture	SNP consequence	HGNC
rs12632456	7.36E-05	AA:2;AG:14;GG:7;00:0	AA:0;AG:10;GG:35;00:1	NON_SYNONYMOUS_CODING	*FLNB*
rs12638356	9.12E-05	GG:3;GA:14;AA:6;00:0	GG:0;GA:12;AA:34;00:0	REGULATORY_REGION	*FLNB*
rs61748935	0.0004	AA:1;AG:5;GG:15;00:2	AA:0;AG:0;GG:42;00:4	NON_SYNONYMOUS_CODING	*C22orf40*
rs8464	0.0005	AA:1;AC:11;CC:11;00:0	AA:0;AC:5;CC:40;00:1	INTERGENIC, REGULATORY_REGION	*NA*
rs8640	0.001	TT:1;TC:14;CC:6;00:2	TT:0;TC:11;CC:28;00:7	SYNONYMOUS_CODING	*FLNB*
rs74207980, rs9270773	0.003	GG:1;GA:12;AA:7;00:3	GG:0;GA:10;AA:28;00:8	INTERGENIC	*NA*
rs2070687	0.004	GG:0;GC:4;CC:19;00:0	GG:1;GC:21;CC:21;00:3	SPLICE_SITE	*SFTPC*
rs35709976	0.004	-/-:0;-/+TA:2;+TA/+TA:21;00:0	-/-:0;-/+TA:19;+TA/+TA:26;00:1	REGULATORY_REGION	*F11*
rs6065, CM032257	0.005	TT:0;TC:11;CC:12;00:0	TT:0;TC:6;CC:37;00:3	NON_SYNONYMOUS_CODING	*GP1BA*
rs7577978	0.005	AA:2;AG:6;GG:8;00:7	AA:0;AG:4;GG:22;00:20	DOWNSTREAM	*ATIC*
rs238239	0.006	CC:3;CT:11;TT:9;00:0	TT:4;TC:27;CC:14;00:1	NON_SYNONYMOUS_CODING	*ENO3*
rs10949870	0.006	AA:3;AG:14;GG:4;00:2	AA:0;AG:16;GG:17;00:13	REGULATORY_REGION	*ZNF727*
rs59337853	0.007	-/-:2;-/T:4/TT:15;00:2	-/-:0;-/T:2;TT:39;00:5	REGULATORY_REGION	*ALOX5*
rs35187157	0.008	-/-:1;-/+TT:8;+TT/+TT:14;00:0	-/-:6;-/+TT:28;+TT/+TT:11;00:1	REGULATORY_REGION	*TRBV30*
rs3787537	0.008	TT:2;TC:11;CC:10;00:0	TT:0;TC:12;CC:32;00:2	REGULATORY_REGION	*SLCO4A1*
rs9332119	0.009	CC:0;CG:10;GG:11;00:2	CC:0;CG:6;GG:35;00:5	WITHIN_NON_CODING_GENE	*CYP2C9*
rs4298	0.009	TT:1;TC:5;CC:13;00:4	TT:0;TC:2;CC:31;00:13	SYNONYMOUS_CODING	*ACE*
rs6958588	0.009	TT:3;TC:13;CC:4;00:3	TT:0;TC:19;CC:17;00:10	WITHIN_NON_CODING_GENE	*ZNF727*
rs976002, rs139555919	0.009	GG:4;GA:10;AA:8;00:1	GG:1;GA:15;AA:27;00:3	NON_SYNONYMOUS_CODING	*TMPRSS11E*
rs2228539	0.009	CC:4;CT:11;TT:8;00:0	CC:2;CT:13;TT:31;00:0	NON_SYNONYMOUS_CODING	*EMR1*
rs11712186	0.0098	CC:1;CT:8;TT:6;00:8	CC:0;CT:7;TT:22;00:17	REGULATORY_REGION	*TOPBP1*

Supplementary Table 4. Top ten Reactome pathways predictive of infectious event.

MCC	AUC (CI%95)	Pathway name	SNPs / genes
0.72	0.88 (0.75 - 0.97)	GPCR downstream signaling	chr2:178528629 (PDE11A), rs4762,CM920009 (AGT), rs78644275 (OR51T1)
0.71	0.89 (0.81 – 0.98)	Bile acid and bile salt metabolism	CM014711 (HSD17B4), rs11045681 (SLCO1B7), rs41272687,CM005424 (CYP27A1)
0.70	0.91 (0.84 – 0.99)	GPCR ligand binding	rs970388 (GABBR2), rs17611 (C5), rs4762,CM920009 (AGT)
0.70	0.91 (0.84 – 0.99)	Peptide ligand-binding receptors	rs2277984 (C3), rs17611 (C5), rs4762,CM920009 (AGT)
		Class A/1 (Rhodopsin-like receptors)	
0.68	0.83 (0.72 – 0.94)	Class I MHC mediated antigen processing & presentation	rs11558955 (RAD23A), rs4036 (TCEB2), rs4673,CM983302 (CYBA)
0.68	0.79 (0.66 – 0.93)	G alpha (q) signalling events	CM012741 (CASR), rs145073237 (XCL1), rs4762,CM920009 (AGT)
0.66	0.89 (0.81 – 0.97)	Response to elevated platelet cytosolic Ca2+	rs216902 (VWF), rs2562830 (TTN), rs6023 (F5)
		Platelet degranulation	
0.62	0.77 (0.63 – 0.90)	G1/S Transition	rs2071467 (TAP2), rs45568137 (PPP2R1B), rs61732929 (POLE)
		Mitotic G1-G1/S phases	
		Mitotic M-M/G1 phases	
0.61	0.77 (0.63 – 0.90)	Gastrin-CREB signalling pathway via PKC and MAPK	CM012741 (CASR), rs148047905 (MMP3), rs4762,CM920009 (AGT)
0.55	0.80 (0.69 – 0.91)	Cytochrome P450 - arranged by substrate type	rs1126545 (CYP2C18), rs58871670 (CYP2B6), rs41272687,CM005424 (CYP27A1)
		Phase 1 - Functionalization of compounds	

Supplementary Table 5. Top ten Reactome pathways predictive of positive/negative culture.

MCC	AUC (CI%95)	Pathway name	SNPs / genes
0.70	0.85 (0.73 – 0.96)	Interferon Signaling	rs1049069 (HLA-DQB1), rs12632456 (FLNB), rs146778723 (HLA-C)
0.62	0.86 (0.77 – 0.95)	Rho GTPase cycle	rs11800462 (TNFRSF25), rs1801284,CM982053 (HMHA1), rs2061821 (AKAP13)
0.60	0.85 (0.75 – 0.95)	G alpha (i) signalling events	rs114642578 (UBD), rs11575580 (IL11RA), rs61745073 (ADCY4)
0.58	0.84 (0.74 – 0.94)	Interleukin-2 signaling	rs11256369 (IL2RA), rs117805308 (CSF2RB), rs290223 (SYK)
		Interleukin-3, 5 and GM-CSF signaling	
		Signaling by Interleukins	
0.53	0.83 (0.72 – 0.93)	Glucose metabolism	rs238239 (ENO3), rs11208257 (PGM1), rs6065,CM032257 (GP1BA)
0.50	0.77 (0.67 – 0.87)	ABCA transporters in lipid homeostasis	rs10491178 (ABCA10), rs1860447 (ABCA9), rs3752232 (ABCA7)
0.45	0.80 (0.69 – 0.90)	Regulation of Lipid Metabolism by Peroxisome proliferator-activated receptor alpha (PPARalpha)	rs20551 (EP300), rs2305160 (NPAS2), rs9627281 (C22orf40, downstream to PPARA)
		PPARA Activates Gene Expression	
		Fatty acid, triacylglycerol, and ketone body metabolism	
0.44	0.78 (0.69 – 0.88)	Platelet Aggregation (Plug Formation)	chr4:155507965 (FGA), rs145155424 (SOS1), rs6065,CM032257 (GP1BA)
0.44	0.76 (0.64 – 0.88)	GPCR ligand binding	CM043760 (F2), rs11575580 (IL11RA), rs16940655 (CRHR1)
0.44	0.69 (0.55 – 0.82)	GPCR downstream signaling	rs11575580 (IL11RA), rs16940655 (CRHR1), rs28371560 (GHRHR)

Paper III

Is not included due to copyright