NGS-PTL Report Summary

Project ID: 306242
Funded under: FP7-HEALTH
Country: Italy

Final Report Summary - NGS-PTL (Next Generation Sequencing platform for targeted Personalized Therapy of Leukemia)

Executive Summary:
Hematological diseases are highly heterogeneous malignancies in terms of molecular mechanisms involved in their onset and progression. Heterogeneity can be further observed within the same hematological disease at the inter-individual level, being reflected by different clinical outcomes and responses to treatment in different patients. Nowadays, the advent of high-throughput next generation sequencing (NGS) technologies is revolutionizing genomics and transcriptomics by providing a single base resolution tool for a unified deep analysis of diseases complexity and allowing a fast and cost-efficient fine-scale assessment of the genetic variability hidden within cohorts of patients affected by the same leukemia type. Therefore, NGS strategies promise to play a crucial role in the selection of tailored therapeutic approaches treatments. The general objectives of NGS-PTL project were the creation of a European Hematological/NGS platform of physicians and scientists aimed at demonstrating the utility of genomic information to improve therapeutic outcomes for therapeutic interventions on different leukemia subtypes and its applicability, as well as of translation of genomic data, essential to predict and personalize medicine applications to the routine clinical practice. This goal was achieved thanks to the development of biostatistics and bioinformatics tools for data analysis and integration of scientific research data with clinical/molecular databases. Results included the evaluation of different tools for fusion genes detection and the development of a WES pipeline for leukemia samples analysis and a combined approach for the detection of driver mutations in WES.

We discovered novel insights into the mechanisms involved in leukemogenesis and we developed genetic models that accurately define novel leukemia subtypes based on the genomic landscape of individual patients, including the single nucleotide variants, expression profiles and fusion transcripts detected via whole exome and whole transcriptome sequencing experiments.

In addition, the consortium identified novel diagnostic, prognostic and minimal residual disease biomarkers by analysing target genes in large patient cohorts. To this purpose, amplicon-based, targeted DNA enrichment-based assays and “leukemia panels” for myeloid and lymphoid malignancies were developed and tested. Ultra-Deep sequencing analysis was used to investigate clonal competition among cell populations harboring mutations in key leukemia genes.

In conclusion, the results of NGS-PTL project represent a step forward to the translation of personalized medicine into clinical practice, which is expected to dramatically impact on real-life of leukemia patients.

Project Context and Objectives:
Nowadays, cancer malignancies account for around 13% of deaths worldwide, showing increasing incidence and impact on human health, thus making the discovery of new therapeutic approaches one of the main medical goals of present-day societies.

The prevalence of haematological diseases especially appears to be higher in elderly populations, so that for industrialized countries, such as the European ones, in which the proportion of elderly people is becoming greater and greater, leukemia-related burden on the health care system is consequently growing.

To date, huge scientific effort was made to describe and treat haematological malignancies, which have been proved to represent a mosaic of disease phenotypes characterized by aberrant genomic programming, especially affecting the structure,
function, and expression patterns of genes involved in key cellular processes, such as proliferation, differentiation, and apoptosis, whose deregulation leads to possible neoplastic transformation. For instance, aberrant gene fusions produced by non-random chromosomal translocations and/or inappropriate expression of oncogenes are currently supposed to represent the main causes of most leukemia subtypes, although de novo or acquired somatic mutations could be essential for disease progression by enabling leukemic cells to become resistant to therapies. Identification of common translocations, as well as of genes located at translocation breakpoints, has led to the development of a series of anti-leukemia treatments able to target specifically mutated loci. Nevertheless, a high number of driver mutations are assumed to be still unknown as far as concerns leukemia subtypes for which the molecular pathogenesis is only partially understood and translation of the available genomic information into biological and clinical consequences is still to be fully achieved.

In fact, although relevant prognostic, and in some cases, therapeutic information have been already obtained, most of these findings have been pointed out by applying candidate-gene or candidate-pathways approaches, which are inherently limited, being focused on the investigation of short lists of preselected genes and variants. This kind of experimental approach has been also proved to be inadequate to effectively explore the heterogeneous landscape of haematological malignancies, which highly differ in terms of both molecular mechanisms related to their development or progression and diversity of phenotypic manifestations. Moreover, a considerable heterogeneity can be further observed within the same hematological disease at the inter-individual level, being reflected by different clinical outcomes and responses to treatment in different patients. According to the vastly different cytogenetic, genetic, and genomic alterations potentially underlying this outstanding heterogeneity, the achievement of accurate molecular classification of hematological diseases appears to be a key milestone in such research field and should play a greater and greater role also in the process of clinical decision making.

For this purpose, the NGS-PTL project has implemented research activities aimed at the investigation of genomic and transcriptomic signatures not only involved in the development of different leukemia subtypes, but also impacting on their clinical outcome and prognosis, thus putting a step forward to the establishment of an interdisciplinary field of molecular medicine able to distinguish between patients who are more likely to benefit from a given therapy and those who are instead bound to not benefit from it or even to experience severe adverse reactions. Therefore, one of the main goals of the project was to overcome the current obstacles towards the translation of personalized medicine into clinical practice by developing and optimizing new experimental and methodological approaches for patients’ stratification. This turned out to be crucial to improve the identification of really effective and not a posteriori potentially dangerous tailored therapeutic treatments. In fact, despite recent advances in the clinical treatment of some leukemia subtypes, several ones continue to have a poor prognosis and, in a proportion of long-term surviving patients, treatment results are unsatisfactory for short and long-term toxicities. As a consequence, early-diagnosis together with specifically tailored approaches still represent key points in determining the health, quality and estimated life of patients, as well as in the sustainability and efficiency of the healthcare system, due to the fact that very expensive, but in some cases useless or even harmful, therapies could be avoided.

To achieve the above-mentioned goal, the project has taken advantage from the outstanding improvements recently observed in the field of massive parallel sequencing technologies that have completely revolutionized genomics and transcriptomics, thus providing a single base resolution tool for a unified deep analysis of the hematological diseases complexity. In particular, by simultaneously evaluating all genes, the recently developed genome-wide approaches based on massive parallel DNA or RNA sequencing were used by the project to depict the as comprehensive as possible picture of the genomic variability hidden within the examined cohorts of leukemia patients, thus providing a nonbiased way to detect the full spectrum of potentially driver mutations and to assess also their cooperativeness in determining disease development of resistance to treatments. The adopted explorative approaches were chosen by the project according to their potential to be clinically relevant in leukemia diagnosis by improving the understanding of functional consequences of sequence and transcription variation. Moreover, these approaches were coupled with the development of analytical methods aimed at integrating all the measured experimental variables into a unique view, in strict relationship with the observed patterns of genomic organization, structure and expression. This required the implementation and optimization of network analyses based on the available a priori biological knowledge and mainly focused on genomic location, biological pathways, interactome and transcriptome complexes. This enabled to go beyond traditional reductionist methods and to point out unexpected relationships not observable with
conventional analyses, thus identifying the higher-scale processes, such as biochemical pathways, or cellular functionalities, that are most involved and perturbed in the hematological diseases under investigation, resulting in a clearer patient stratification or in a better understanding of the origin of different responses to therapeutic interventions.

Accordingly, the experimental and analytical framework of the project succeeded in enabling both pre-clinic and diagnostics studies characterized by a cost-effective depiction of a comprehensive catalogue of diagnostic and prognostic markers able to guide the targeting of therapeutic interventions.

This enabled to achieve an actual objectification of the concept of personalized medicine by allowing pragmatic exploitation of information emerged from the as full as possible characterization of each patient’s genomic and transcriptomic profile into the processes of patient groups stratification that are necessary to provide guidance to several hematological diseases therapeutic interventions.

However, the progressive shift from a genetic to a genomic perspective that is occurring also in the field of molecular medicine entails the necessity to establish interdisciplinary research networks able to include also highly specialized biotechnological and bioinformatics expertise and facilities. In fact, large-scale parallelization of third generation sequencing platforms results in billions of sequence reads from both whole exome and whole transcriptome experiments that have to be computationally filtered for quality check and assembled. Due to this issue, computational power for pre-processing of the generated raw data and ensured by large CPU clusters, as well as for downstream bioinformatics analyses have become of primary importance to manage, analyse, and interpret this overwhelming amount of data in order to draw meaningful conclusions. This necessity was fulfilled during the project by the establishment of a European Hematological/NGS Platform assembled to represent a network of scientists focused on the implementation of massive parallel sequencing approaches to the study of a variety of hematological malignancies and, in particular, of different leukemia subtypes. The constitution of such network enabled the project to benefit from the synergistic value offered by different clinical research groups, as well as by highly qualified research partners with specific expertise in the field of massive parallel sequencing and bioinformatics.

Accordingly, partners characterized by longstanding experience in pre-clinical and clinical studies of different leukemia subtypes made available to the project large cohorts of well-characterized leukemia samples coupled with the related clinical and/or biological data. Research units with informatics and data management expertise instead contributed to the creation of a dedicated web-based informatics platforms aimed at storing and organizing all biological and clinical data, as well as the results of experimental procedures carried out during the whole project, to effectively share data both among clinical research groups and between them and the biotechnological ones. Finally, partners characterized by highly specialized biotechnological and bioinformatics backgrounds provided expertise and facilities essential to design and implement the scheduled massive parallel sequencing approaches, as well as to develop and optimize ad hoc bioinformatics pipelines for the analysis of the generated data.

The project thus succeeded in the pragmatic effort of organizing several international research groups into a structured collaborative network that was able to realize a series of incisive pre-clinical and clinical studies on different leukemia subtypes, all characterized by the exploitation of recent and innovative advances in the field of DNA and RNA massive parallel sequencing for the depiction of an exhaustive picture of the leukemia genome complexity.

Accordingly, new research, diagnostic and prognostic approaches to the study of several leukemia subtypes were identified and validated by project activities, leading to the achievement of the following specific objectives:

- Development of a European Hematological/NGS network of physicians and scientists aimed at demonstrating the utility of genomic information to improve the outcomes for therapeutic interventions on different leukemia subtypes, as well as of translation of genomic data, essential to predict and personalize medicine applications to routine clinical practice.

- Setting up and sharing among the network partners of optimized protocols for collection, storage and processing of biological samples and genomic data from leukemia patients previously (retrospectively) and prospectively enrolled in National and European clinical trials with conventional chemotherapy, monoclonal antibodies or tyrosine kinase inhibitors (TKIs).

- Discovery of novel insights into the mechanisms involved in leukemogenesis and development of genetic models that
accurately define novel leukemia subtypes based on the genomic landscape of individual patients, and especially on the single nucleotide variants, expression profiles and fusion transcripts detected via whole exome and whole transcriptome sequencing experiments.

- Development of biostatistics and bioinformatics tools for coupling scientific research data with clinical/molecular databases to ensure patient specific early diagnosis and prediction of treatments' sensitivity and adverse effects though the evaluation of the impact of their genomic signatures on risk-assessment and clinical outcome.

- Fine distinction between patients who are more likely to benefit from a given therapy and those who are predicted to experience severe adverse reactions in order to identify really effective and not a posteriori potentially dangerous tailored therapeutic treatments, especially for elderly patients, which can be specifically directed only to individuals for which they are safe and useful.

- Setting up of amplicon-based and/or targeted DNA enrichment-based assays to identify novel diagnostic, prognostic and minimal residual disease follow-up biomarkers, as well as of ultra-deep sequencing protocols for evaluation of clonal competition among cell populations harboring mutations in key leukemia genes with respect to their selection by therapy and with regard to the presence of respective sub clones at the time of diagnosis.

- Identification of transcription signatures allowing the stratification of syndromes that might evolve in tumors to detect signatures that might be predictive of disease evolution, as well as to point out molecular features that allow stratification of certain leukemias in subgroups on the basis of clinical features, such as response to treatment and prognosis.

- Selection of genomic alterations with a potential role as diagnostic and prognostic biomarkers for the examined leukemias among those identified with whole exome and whole transcriptome sequencing experiments, to develop “leukemia diagnostic panels” of specific mutations, genomic/transcript alterations and miRNA biomarkers, to enable fast and comprehensive diagnostic screening, analysis of risk-assessment and prediction of therapy-failure, as well as to include relevant genomic alterations in novel clinical trials to target therapies for a more effective cure of the disease, with a lower impact in terms of secondary effects on the patient.

- Communication and transfer of the knowledge generated within the project to the scientific, commercial, policy, and general audience, as well as construction of a data warehouse that will allow the research community to access the data generated within the project.

Project Results:
Introduction
The activities performed within the framework of the NGS-PTL project led to the publication of 40 peer-reviewed papers and to the submission of 3 patents on miRNA signature in leukemia patients. Moreover, NGS-PTL partners have presented the results of the consortium projects in several international events. Accordingly, dissemination activities included more than 150 poster presentations, oral communications, seminars, lectures, webinars and other activities.

The scaffold on which the consortium built its main achievements in terms of scientific and technological results comprised:
- standardized protocols for different NGS approaches;
- bioinformatics pipelines for NGS data analysis;
- more than 400 leukemia cases sequenced;
- an output of raw data exceeding 10 TB;
- a secure database to store patient confidential data.

Here, we are going to highlight the main results obtained within the NGS-PTL activities.

Bioinformatics and biostatistics tools for NGS data analysis
Whole-exome sequencing
The consortium developed optimized tools for variant calling and prioritization in leukemia samples.

Optimized pipeline of MuTect and GATK tools to improve the detection of somatic single nucleotide polymorphisms in whole-exome sequencing data
Detecting somatic mutations in whole exome sequencing data of cancer samples has become a popular approach for profiling cancer development, progression and chemotherapy resistance. Several studies have proposed software packages, filters and parametrizations. However, many research groups reported low concordance among different methods. We aimed to develop a pipeline that detects a wide profile of single nucleotide mutations with high validation rates. We bound together two standard softwares – Genome Analysis Toolkit (GATK) and MuTect – and adapted their algorithms to create the GATK-LODN method. As proof of principle, we applied to our pipeline exome sequencing samples of hematological (Acute Myeloid and Acute Lymphoblastic Leukemias) and solid (Gastrointestinal Stromal Tumor and Lung Adenocarcinoma) tumors. We performed experiments on simulated data to test the sensitivity and specificity of our pipeline.
MuTect presented the highest validation rate (90%), but limited number of somatic mutations detected. The GATK detected a high number of mutations but with low specificity. The GATK-LODN increased the performance of GATK results (from 37 to 70% of confirmed variants), while preserving important mutations not detected by MuTect. However, GATK-LODN filtered more variants in the hematological samples than in the solid tumors. Experiments in simulated data demonstrated that GATK-LODN increased both specificity and sensitivity of GATK results.
We presented a pipeline that detects a wide range of somatic single nucleotide polymorphisms, with good validation rates, from exome sequencing data of cancer samples. We also showed the advantage of combining standard algorithms to create the GATK-LODN method that increased specificity and sensitivity of GATK results. This pipeline can be helpful in discovery studies aimed to profile the somatic mutational landscape of cancer genomes.

Functional interpretation of variant and non-variant positions in whole-exome sequencing data
Methods for the analysis of whole-genome and -exome data are typically based on detection of variant sequences by comparison of sequencing reads with the reference human genome which are then stored in the Variant Call Format (VCF). However, the reference sequence contains both common and rare disease risk variants, including rare susceptibility variants for acute lymphoblastic leukemia and the Factor V Leiden allele associated with hereditary thrombophilia [Chen et al. 2011]. In fact, out of 16,400 variant positions associated with disease, more than 4,000 variants are represented in the reference genome by the minor allele [Dewey et al. 2011]. This poses a serious limitation to a comprehensive evaluation and detection of markers related for example to response to drugs and treatments. To overcome this problem were developed, in collaboration with Knome Inc., an integrated hardware-software platform for the annotation and interpretation of variation in genomes/exomes. The platform uses the emerging genome Variant Call Format (gVCF) which allows to store the information of both variant and non-variant positions in the genome/exome. Analysis of patients with the pipeline based on knome system (knoSYS 100) identified about 500 variants represented in the genome by a minor allele (MAF less than 5%) per patient. These variants was correlated with clinical data as they may represent important candidates as molecular markers of predisposing factors of disease and of response to drugs.

A specific “leukemia diagnostic panels” based on selection of novel relevant biomarkers for leukemia diagnosis, prognosis and therapeutic decision making
The identification of a biomarker from NGS is coupled with statistical and functional analysis aimed to identify genes carrying potential driver mutations. Somatic mutations identification step was setup with a combined approach that integrates GATK software for calling of Indels and MuTect for calling of SNPs which allowed to achieve a 89% validation rate. Different strategies can be employed to detect putative genes carrying driver mutations based for example on the not-random distribution of driver mutations in the genome or on the predicted effect of mutations. To maximize the sensitivity of the detection of drivers to be used as candidate leukemia biomarkers, a combined approach of three complementary statistical methods was chosen. The first method identifies genes significantly enriched in somatic mutations (MutSigCV), the second is
based on mutation clustering on protein domains (OncodriveClust) while the third identifies genes accumulating putative driver mutations with high functional impact (OncodriveFM). Finally candidate driver genes detected by one or more methods are mapped on a functional network to identify driver genes perturbing the same metabolic pathway or functional module. Using the method setup were analyzed the sequencing data generated from cohorts of patients affected by 4 different leukemia types (AML, ALL, CLL and ET) detecting 47 putative biomarkers which was used as a base to build diagnostic kits.

Whole transcriptome sequencing
Recently, many tools for chimera identification were reported. The consortium tested and compared several of them and established a workflow for WTS data analysis, including detection of fusion genes and alternative splicing events.

State-of-the-art fusion-finder algorithms sensitivity and specificity
We tested eight fusion-detection tools (FusionHunter, FusionMap, FusionFinder, MapSplice, deFuse, Bellerophontes, ChimeraScan, and TopHat-fusion) to detect fusion events using synthetic and real datasets encompassing chimeras. The comparison analysis run only on synthetic data could generate misleading results since we found no counterpart on real dataset. Furthermore, most tools report a very high number of false positive chimeras. In particular, the most sensitive tool, ChimeraScan, reports a large number of false positives that we were able to significantly reduce by devising and applying two filters to remove fusions not supported by fusion junction-spanning reads or encompassing large intronic regions. The discordant results obtained using synthetic and real datasets suggest that synthetic datasets encompassing fusion events may not fully catch the complexity of RNA-seq experiment. Moreover, fusion detection tools are still limited in sensitivity or specificity; thus, there is space for further improvement in the fusion-finder algorithms.

Chimera: a Bioconductor package for secondary analysis of fusion products
Chimera is a Bioconductor package that organizes, annotates, analyses and validates fusions reported by different fusion detection tools; current implementation can deal with output from bellerophontes, chimeraScan, deFuse, fusionCatcher, FusionFinder, FusionHunter, FusionMap, mapSplice, Rsubread, tophat-fusion and STAR. The core of Chimera is a fusion data structure that can store fusion events detected with any of the aforementioned tools. Fusions are then easily manipulated with standard R functions or through the set of functionalities specifically developed in Chimera with the aim of supporting the user in managing fusions and discriminating false-positive results.

Alternative splicing detection workflow needs a careful combination of sample prep and bioinformatics analysis
RNA-Seq provides remarkable power in the area of biomarkers discovery and disease characterization. Two crucial steps that affect RNA-Seq experiment results are Library Sample Preparation (LSP) and Bioinformatics Analysis (BA). We describes an evaluation of the combined effect of LSP methods and BA tools in the detection of splice variants. Different LSPs (TruSeq unstranded/stranded, ScriptSeq, NuGEN) allowed the detection of a large common set of splice variants. However, each LSP also detected a small set of unique transcripts that are characterized by a low coverage and/or FPKM. This effect was particularly evident using the low input RNA NuGEN v2 protocol. A benchmark dataset, in which synthetic reads as well as reads generated from standard (Illumina TruSeq 100) and low input (NuGEN) LSPs were spiked-in was used to evaluate the effect of LSP on the statistical detection of alternative splicing events (AltDE). Statistical detection of AltDE was done using as prototypes for splice variant-quantification Cuffdiff2 and RSEM-EBSeq. As prototype for exon-level analysis DEXSeq was used. Exon-level analysis performed slightly better than splice variant-quantification approaches, although at most only 50% of the spiked-in transcripts was detected. The performances of both splice variant-quantification and exon-level analysis improved when raising the number of input reads. Data, derived from NuGEN v2, were not the ideal input for AltDE, especially when the exon-level approach was used. We observed that both splice variant-quantification and exon-level analysis performances were strongly dependent on the number of input reads. Moreover, the ribosomal RNA depletion protocol was less sensitive in detecting splicing variants, possibly due to the significant percentage of the reads mapping to non-coding transcripts.

Identification of novel mutations and molecular profiles by exome and/or transcriptome NGS

Page 6 of 26
Research and Innovation
Acute Myeloid Leukemia

Acute Myeloid Leukemia (AML) is a highly heterogeneous disease and a complex network of events contribute to its pathogenesis. Recently, functional categorization of mutated genes in AML identified 9 classes of affected genes [The Cancer Genome Atlas Research Network, 2013]. However, how genomic alterations cooperate to induce AML, the pathways affected by the mutated genes and their prognostic value are still unknown. The main inclusion criteria for patients’ enrollment in the NGS-PTL project was the presence of chromosomal instability in terms of whole or partial chromosome number aberrations and karyotypic complexity.

Genome-wide and Exome analysis of alterations in Acute Myeloid Leukemia patients

We characterized 279 AML patients at diagnosis by SNP Array 6.0 or Cytoscan HD Array (Affymetrix), in order to detect copy number aberrations (CNAs). Thirty-three samples were also analyzed by Whole Exome Sequencing WES (HiSeq 1000, Illumina).

To explore mechanisms involved in the mutational processes in which cancer cells have been involved, we generated catalogs of somatic mutation from our AML cohort and applied mathematical methods to extract mutational signatures of the underlying processes. Six mutational signatures were identified, all characterized by prominence of C>T substitutions at NpCpG sites, suggesting a process in which 5-methyl-Cytosine are subject to spontaneous deamination, which has been correlated to ageing [Alexandrov et al. 2013].

To better stratify AML patients and identify novel molecular biomarkers, we mapped genomic alterations (single nucleotide variants, SNVs, and CNAs) in biological pathways and correlated it with clinical outcome.

We mapped SNVs in biological pathways using KEGG database and we identified putative driver genes with DOTS-finder. KEGG Pathways signatures analysis was able to stratify our cohort in 3 groups. We focused on the signatures of each group and we highlighted the biologically relevant pathways and potential biomarkers, according to WES data. We were able to distinguish 3 groups of patients: two of them were characterized by distinct molecular biomarkers (KRAS and TP53) able to stratify patients with different prognosis. Moreover, KRAS was identified as putative driver gene in AML by two different tools, underling its importance in the disease pathogenesis. On the other hand, we were not able to identify a unique molecular marker that characterizes the third group. However, an enrichment of mutated genes involved in several metabolic pathways was detected.

We detected CNAs involving JAK2, which may have a role in overall survival rate. Moreover, the profile of deleted regions was assessed and we combined the deletome profile with WES data. Interestingly, we found deletions of genes that are also targeted by mutations (such as BRCA2, LRRK1). Moreover, these genes were involved in pathways affected by genomic mutations (i.e. CASK deletion and MPP6 point mutation, CDK6 deletion and PPM1B point mutation, MAPT deletion and SPAG5 point mutation).

By SNP array we have identified CNAs involving novel potential leukemia-related genes. Our results suggest that the comparison between SNP and WES data could provide important findings on the prognosis of AML patients. Minimal deleted regions deserve further investigation in order to identify new candidate oncogenes which could be relevant AML biomarkers.

A Specific Pattern of Somatic Mutations Associates with Poor Prognosis Aneuploid Acute Myeloid Leukemia

One of the inclusion criteria to enroll AML patients in the NGS-PTL project, was the presence of chromosomal aberrations at the time of diagnosis/relapse, in order to identify AML-specific alterations having a causative and/or tolerogenic role towards aneuploidy. To this purpose, we performed whole exome sequencing (WES, Illumina Hiseq1000) of 70 samples from our Aneuploid-AML (A-AML) and Euploid-AML (E-AML) cohort. Gene expression profiling (GEP) and SNP array was also performed.

We detected a significantly higher mutation load in A-AML compared with E-AML (median number of variants: 31 and 15, p=.04) which was interestingly unrelated to patients’ age (median age: 63.5 years in A-AML and 62 years in E-AML, Xie et al, Nat. Med. 2014).

WES analysis also revealed a specific pattern of somatic mutations in A-AML. A-AML had a lower number of mutations in signaling genes (p=.04), while being enriched for alterations in cell cycle genes (p=.01) compared with E-AML. The mutated genes were involved in different cell cycle phases, including DNA replication, centrosome dynamics, chromosome segregation,
mitotic checkpoint and regulation. Moreover, genomic deletion of cell cycle-related genes was frequently detected in A-AML. Notably, ESPL1 which associated with aneuploidy, chromosome instability and DNA damage in mammary tumors [Mukherjee et al. Oncogene 2014] was mutated and also upregulated in A-AML compared with E-AML (p=.01), the latter showing expression levels comparable to controls. Among the top-ranked genes differentially expressed between A-AML and E-AML, we identified a specific signature, which has been previously linked to defects in chromosome number. Additional mutations targeting DNA damage and repair pathways were identified in A-AML, including TP53 mutations, which account for 15% of cases. Moreover, A-AML showed a significant upregulation of a KRAS transcriptional signature and downregulation of FANCL- and TP53-related signatures, irrespective of TP53 mutational status.

Our data show a link between aneuploidy and genomic instability in AML. Deregulation of the cell cycle machinery, DNA damage and repair checkpoints either through mutations, copy number and transcriptomic alterations is a hallmark of A-AML. The results define specific genomic and transcriptomic signatures that cooperate with leukemogenic pathways, as KRAS signaling, to the development of the aggressive phenotype of A-AML and suggest that a number of A-AML patients may benefit from pharmacological reactivation of TP53 pathway (e.g. MDM2 inhibitor, clinical trial NP28679).

Mutational landscape of del(9q) Acute Myeloid Leukemia

A cohort of 20 paired-cases with a deletion of the long arm of chromosome 9 were selected. Deletion of the long arm of chromosome 9, del(9q), is a recurrent genomic abnormality, which occurs at a frequency of ~2% in AML. Interestingly, deletions of 9q are mainly found in t(8;21)-positive AML, as well as in AML with NPM1 (NPM1mut) or CEBPA (CEBPAmut) gene mutation, thereby suggesting that del(9q) can act as cooperating event in these prognostically favorable AML subgroups. Samples were previously characterized by SNP 6.0 microarray analysis to delineate the minimally deleted region (MDR) on 9q encompassing seven genes (GKAP1, KIF27, C9orf64, HNRNPK, RM1, SLC28A3, NTRK2). Moreover, by targeted resequencing in n=50 non-9q deleted cases, a mutation in HNRNPK, recently confirmed to be recurrently mutated by The Cancer Genome Atlas (TCGA) project, was detected. These findings pointed to HNRNPK as the most important candidate gene of the MDR. To further evaluate the biology underlying 9q deleted/HNRNPK haploinsufficient cases, gene expression data were generated by microarray technology comparing NPM1mut cases with and without del(9q) (n=11 vs n=119, respectively). These analyses showed deregulated expression of genes involved in splicing and mRNA processing, and there was an overlap with gene expression changes following shRNA-mediated HNRNPK knock-down in AML cell lines, which also suggested a growth advantage for haploinsufficient cells. While these data further support that HNRNPK might play a cooperating role in AML, WES is employed to see whether there are additional mutations commonly linked to del(9q). By WES, were detected on average 7.8 somatic protein altering point mutations per sample (missense and nonsense SNVs) and 2.5 frameshift insertions or deletions affecting genes known to play a role in AML as well as genes not yet linked to AML. In accordance with the general mutational spectrum of t(8;21), NPM1 or CEBPA mutant AML, were identified mutations in known epigenetic regulators such as ASXL1, ASXL2, TET2 or DNMT3A, but was also find novel somatic mutations in additional genes involved in the regulation of the chromatin structure such as BRD3 or BRWD3. Furthermore, were identified mutations in genes associated with mRNA processing and RNA splicing, as well as mutations affecting the RAS-signaling pathway and DNA repair mechanisms.

RNA Sequencing reveals novel and rare fusion transcripts in Acute Myeloid Leukemia

It has been reported that chromosomal rearrangements and fusion genes have a crucial diagnostic, prognostic and therapeutic role in AML. A recent RNA sequencing (RNAseq) study on 179 AML revealed that fusion events occur in 45% of patients [The Cancer Genome Atlas Research Network, 2013]. However, the leukemogenic potential of these fusions and their prognostic role are still unknown.

To identify novel rare gene fusions having a causative role in leukemogenesis and to identify potential targets for personalized therapies, transcriptome profiling was performed on AML cases with rare and poorly described chromosomal translocations. Bone marrow samples were collected from 5 AML patients (#59810, #20 and #84 at diagnosis and #21 and #32 at relapse). RNAseq was performed using the Illuma Hiseq1000 platform. The presence of gene fusions was assessed with deFuse and Chimerascan. Putative fusion genes were prioritized using Pegasus and Oncofuse, in order to select biologically relevant fusions. The fusions were prioritized according to mapping of partner genes to chromosomes involved in the translocation or to Chimerascan and deFuse concordance.
The CBFβ-MYH11 chimera was identified in sample #84, carrying inv(16) aberration, thus confirming the reliability of our analysis.

Sample #59810 carried the fusion transcript ZEB2-BCL11B (Driver Score, DS=0.7) which is an in-frame fusion and a rare event in AML associated with t(2;14)(q21;q32). The breakpoint of the fusion mapped in exon 2 of ZEB2 and exon 2 of BCL11B. Differently from previous data [Torkildsen, et al, 2015], this fusion transcript showed 3 splicing isoforms. Type 1 isoform is the full-length chimera and it retains all exons of both genes involved in the translocation. Type 2 isoform was characterized by the junction of exon 2 of ZEB2 and exon 3 of BCL11B. In type 3 isoform, exon 2 and 3 of BCL11B were removed, resulting in an mRNA composed by exon 2 of ZEB2 and exon 4 of BCL11B. Gene expression profiling showed an upregulation of ZEB2 and BCL11B transcripts in the patient's blasts, compared to 53 AML samples with no chromosomal aberrations in the 14q32 region. The same samples showed the WT1-CNOT2 chimera, which is a novel out-of-frame fusion (DS= 0.008) related to t(11;12) translocation, identified by cytogenetic analysis.

Two new in-frame fusion genes were identified in sample #20: CPD-PXT1 (DS= 0.07) which appeared as the reciprocal fusion product of t(6;17) translocation, and SAV1-GYPB, which remained cryptic at cytogenetic analysis (DS= 0.8 alternative splicing events are being investigated). SAV1 was downregulated in sample #20 compared to our AML cohort, suggesting the putative loss of a tumour-suppressor gene.

Sample #21 carried a t(3;12) translocation and RNAseq identified a novel fusion event between chromosomes 19 and 7, involving the genes OAZ and MAFK (DS= 0.9). Finally, no chimeras were confirmed in sample #32 having a t(12;18) translocation.

Our data suggest that fusion events are frequent in AML and a number of them cannot be detected by current cytogenetic analyses. Gene fusions cooperate to AML pathogenesis and heterogeneity and we are further investigating the oncogenic potential of the identified translocations. Moreover, the results firmly indicate that different approaches, including G-banding, molecular biology, bioinformatics and statistics, need to be integrated in order to better understand AML pathogenesis and improve patients’ stratification. High-resolution sequencing analysis currently represent the most informative strategy to tailor personalized therapies.

Acute Lymphoblastic Leukemia

High-resolution genome-wide profiling analysis of B-cell precursor acute lymphoblastic leukemia (BCP-ALL) samples previously identified many novel somatic genetic alterations, several of which have clear implications for risk stratification or future therapeutic targeting. However, most of the studies focused on pediatric cases. Therefore a deep molecular characterization of adult patients is still challenging, especially for those cases lacking recurrent fusion genes. Moreover, T-ALL are caused by a combination of fusion genes, over-expression of transcription factors and cooperative point mutations in oncogenes and tumor suppressor genes, which requires deep investigation.

Clustering Adult Acute Lymphoblastic Leukemia (ALL) Philadelphia Negative (Ph-) by Whole Exome Sequencing (WES) analysis

We performed whole exome experiments (Illumina Hiseq2000) to discover novel insights into the mechanisms involved in leukemogenesis and to develop genetic models that accurately define novel adult Ph-negative B-ALL subtypes (and negative for the recurrent known molecular rearrangements) based on the genomic profile of individual adult patients. Point mutations are the prevalent mechanism identified in our cohort (41 patients; 75.5% of SNVs) and Indels are less represented (21.5%). Analysis of SNVs confirmed mutations in important genes known to be involved in leukemogenesis (PAX5, JAK2, TP53, PTPN11). Using KEGG database we mapped the 651 mutated genes to detect the mostly represented pathways. The Jak-STAT signaling pathway (11 genes) and the Cell Cycle (13) pathways resulted to be significantly enriched in our cohort, which may be effectively targeted by currently available JAK inhibitors. We further investigate through an amplicon-based approach the molecular markers identified in large Ph -negative B-ALL cohorts. Preliminary results showed that TP53 and PTPN11 are the most frequently mutated genes (25% and 10%, respectively) in this subtype of B-ALL leukemia.

Comprehensive analysis of Transcriptome variation uncovers known and novel driver events in T-Cell Acute Lymphoblastic Leukemia

We analyzed 31 T-ALL patient samples and 18 T-ALL cell lines by high-coverage paired-end RNA-seq. First, we optimized the
detection of SNVs in RNA-seq data by comparing the results with exome re-sequencing data. We identified known driver genes with recurrent protein altering variations, as well as several new candidates including H3F3A, PTK2B, and STAT5B. Next, we determined accurate gene expression levels from the RNA-seq data through normalizations and batch effect removal, and used these to classify patients into T-ALL subtypes. Finally, we detected gene fusions, of which several can explain the over-expression of key driver genes such as TLX1, PLAG1, LMO1, or NXX2-1; and others result in novel fusion transcripts encoding activated kinases (SSBP2-FER and TPM3-JAK2) or involving MLLT10. In conclusion, we present novel analysis pipelines for variant calling, variant filtering, and expression normalization on RNA-seq data, and successfully applied these for the detection of translocations, point mutations, INDELs, exon-skipping events, and expression perturbations in T-ALL.

Systemic Mastocytosis
According to the World Health Organization (WHO) classification, the diagnosis of Systemic Mastocytosis (SM) relies on bone marrow (BM) examination and is based on a major and four minor criteria. The somatic 'autoactivating' point mutation D816V in the KIT receptor gene is one of the minor criteria, founded in the great majority of patients (90%) and it plays a central role in the pathogenesis of the disease. Nevertheless, morphological and clinical diversity, as well as the fact that some patients are negative for KIT mutations, suggest that the underlying molecular picture is far from being fully elucidated. To shed further light on this issue, we undertook an integrated molecular genetic study with an integrated molecular characterization study of ASM and MCL to identify novel, functionally relevant molecular lesions and/or clinically actionable signaling pathways.

Genome-Wide Molecular Portrait of Aggressive Systemic Mastocytosis and Mast Cell Leukemia Depicted By Whole Exome Sequencing and Copy Number Variation Analysis
A discovery panel including 6 patients with ASM and 6 patients with MCL was studied using whole exome sequencing (WES) and copy number variation (CNV) analysis. Genes were selected for further assessment when recurrently mutated in ≥2 patients or concurrently identified in WES and CNV analyses or previously associated with leukemogenesis or cancer pathogenesis. Among these, genes already reported to be affected by mutations in SM included TET2, NRAS, ASXL1, CBL, IDH1, SRSF2, SF3B1, RUNX1. We also identified genetic alterations in genes not previously implicated in SM pathogenesis including TP53BP1, RUNX3, NCOA2, CDC27, CCND3, E1F2, MLL3, ARID1B, ARID3B, ARID4A, SETD1A, SETD1B, KDM1B, PRDM1, ATM, WRN. A long tail of infrequently mutated genes dominated, resulting in significant intertumoural heterogeneity. However, when genes were assigned to functional pathways to discern patterns of mutations across different patients, we found that PI3K/Akt and MAPK pathways, calcium pathway, chromatin modification, DNA methylation, and DNA damage repair were consistently affected. Interestingly, two loss-of-function mutations (a nonsense and a frameshift mutation) inactivating both alleles of the SETD2 gene were identified in a KIT gene mutation-negative MCL case, who came to our attention in 2012.

WES and CNV analyses of ASM and MCL revealed a complex landscape, not unexpected when considering the clinical heterogeneity of these patients. Nonetheless, key pathways were found to be recurrently altered. Further investigation of selected candidate genes and pathways is warranted and will cast light on the cooperative genetic (and epigenetic?) events underlying the more aggressive forms of SM – paving the way to a better prognostic stratification and more effective treatment.

Essential Thrombocytemia
The JAK2 p.V617F MP p.W515K/L and CALR indels occur in a mutually exclusive pattern in 80-90% of Essential Thrombocythemia (ET) cases. However, the driver mutations are unknown in the remaining 10-20% of cases. We aimed to identify driver mutations in the group of triple negative (TN) ET by exome sequencing.

MPL p.S204P is a recurrent Mutation in Essential Thrombocytemia
We found 27 somatic variants, including indels, in 6 out of 10 TN ET patients (range: 1-10 mutations/case; mean: 2.7 mutations/case), none of which were recurrent. In one case, we found a MPL p.S204P mutation, which is located in the extracellular domain of the MPL receptor. By Sanger sequencing of MPL exon 4 in 20 additional TN ET cases, an additional patient with the MPL S204P mutation was identified.
In order to study the effect of this mutation on the function of MPL, we produced stable Ba/F3 cell lines expressing MPL S204P, MPL W515K or MPL WT, and assessed the dependence of their growth on exogenous thrombopoietin (TPO). Using flow cytometry, we also explored cell surface marker expression on peripheral blood platelets from the two MPL S204P ET patients. Data were compared with healthy donors or ET patients with JAK2 or CALR mutations. In addition, there was a trend for higher expression of KIT, CD36 and CD42b on platelets from the MPL S204P ET cases. Moreover, following platelet activation through the protease activated receptor 1, the degranulation response of platelets from MPL S204P ET was decreased in comparison with JAK2 or CALR mutated ET. The MPL S204P mutation is a recurrent mutation in TN ET, with a frequency of 7% (2/30) in this series, but this mutation does not induce TPO-independent growth nor increased TPO-sensitivity in Ba/F3 cells. However, preliminary phenotypic and functional evidence supports the notion that MPL S204P platelets display specific characteristics as compared with JAK2 or CALR mutated ET. The mechanisms by which the MPL S204P mutation influences megakaryopoiesis and platelet function remain to be elucidated.

Chronic Lymphocytic Leukemia
The high-throughput analysis of CLL cases focused on two different cohorts: the first included patients gaining cytogenetic abnormalities during the progression of the disease, the second comprised patients without any aberration in the TP53 gene at the time of diagnosis. To drive the interpretation of putative somatic variants identified in the first cohort of patients, mutations were annotated using the 1000 Genomes, ESP6500, dbSNP, COSMIC and ClinVar known variants and mutations data. Moreover, several known CLL-related genes were identified (MYD88 p less than 0.0081 KLHL6 p less than 0.0157 SF3B1 less than 0.0685 and NOTCH1 p less than 0.0699). Regarding the second cohort, we focused on the possible genetic markers connected with the selection of TP53 gene defects after therapy as this abnormality is associated with very poor prognosis of affected patients. Moreover, to ascertain the molecular consequences of complex genomic rearrangements, we have employed whole transcriptome sequencing in 12 CLL patients with either chromothripsis (10 patients) or jumping translocations (2 patients).

Amplicon-based and targeted sequencing approaches
In order to deeply investigate the contribution of oncogenic events to leukemia onset and progression and the clonal architecture of leukemia samples, amplicon based and targeted sequencing approaches were set up by NGS-PTL partners, which included the analysis of the most relevant oncogenes and tumor suppressor genes and the correlation with disease stage, clinical outcome and genomic profile. Moreover, ultra-deep sequencing protocols for evaluation of clonal competition among cell populations harboring mutations in key leukemia genes were set up.

Acute Myeloid Leukemia
TP53 mutations are mutually exclusive with FLT3 and NPM mutations in AML patients and are strongly associated with complex karyotype and poor outcome
To further investigate the role and the frequency of TP53 mutations in adult AML, the types of mutations, the associations with recurrent cytogenetic abnormalities and their relationship with response to therapy, clinical outcome and finally their prognostic role, 172 adult AML patients were examined for TP53 mutations using several methods, including Sanger sequencing, Next-Generation Deep-Sequencing (Roche) and HiSeq 2000 (Illumina) platform. 40 samples were genotyped with Genome-Wide Human SNP 6.0 arrays or with CytoScan HD Array (Affymetrix) and analyzed by Nexus Copy Number v7.5 (BioDiscovery). TP53 is the most frequently mutated gene in human tumours. TP53 mutation rate in AML was reported to be low (2.1%), but the incidence of TP53 mutations in AML with a complex aberrant karyotype is still debated. Our cohort was characterized by a median age of TP53 mutated and wild type patients of 68 years (range 42-86), and 65 years (range 22-97) respectively. Conventional cytogenetics showed that:
 a) 52 patients (30.2%) had 3 or more chromosome abnormalities, i.e. complex karyotype;
b) 71 (41.3%) presented with one or two cytogenetic abnormalities (other-AML);
c) 34 patients (19.8%) had normal karyotype.
Most of the TP53 mutated patients (79.3%) had complex karyotype, whereas only 6/29 mutated patients had “no complex karyotype” (21% and 3% of the entire screened population, respectively). Overall, TP53 frequency was 44.2% in the complex karyotype group, suggesting a pathogenetic role of TP53 mutations in this subgroup of leukemia. As far as the types of TP53 alterations regards, the majority of mutations (32) were deleterious.

Copy Number Alterations (CNAs) analysis performed on 40 cases by Affymetrix SNP arrays showed the presence of several CNAs in all cases: they ranged from loss or gain of the full chromosome (chr) arm to focal deletions and gains targeting one or few genes involving macroscopic (>1.5 Mbps), submicroscopic genomic intervals (50 Kbps - 1.5 Mbps) and LOH (>5 Mbps) events. Of relevance, gains located on chr 8 were statistically associated with TP53 mutations (p = 0.001). In addition to the trisomy of the chr 8, others CNAs, located on chromosomes 5q, 3, 12, 17 are significantly associated (p = 0.05) with TP53 mutations. WES analysis was performed in 37 patients: 32 TP53 were wilde-type, while 5 patients were TP53 mutated. Interestingly, TP53 mutated patients had more incidence of complex karyotype, more aneuploidy state, more number of somatic mutations (median mutation rate 30/case vs 10/case, respectively).

Moreover, we investigated the correlation of TP53 mutational status with known molecular alterations (FLT3 and NPM) and we detected a mutually exclusive association between FLT3 and/or NPM1 mutations.

Regarding the clinical outcome, as previously reported [Grossmann V. et Al. Blood 2013]) alterations of TP53 were significantly associated with poor outcome in terms of both overall survival (median survival: 4 and 31 months in TP53 mutated and wild type patients, respectively; p less than 0.0001) and relapse free-survival (RFS) (p less than 0.0001). For these reasons, TP53 mutation screening should be recommended at least in CK-AML patients.

Detection of FLT3 ITD mutated clones by ultra-deep sequencing analysis has important clinical implications in AML patients
FLT3 internal tandem duplication (ITD), one of the most frequent mutations in AML, is reported to be an unstable marker, as it can evolve from FLT3 ITD- to ITD+ during the disease course. Moreover, several TKIs are being tested in AML; therefore, the mutational status of FLT3 may represent an essential criterion for the enrollment of patients into these trials. Accordingly, we developed an amplicon-based ultra-deep-sequencing (UDS) approach for FLT3 mutational screening. We exploited this highly sensitive technology for the retrospective screening of diagnosis, relapse and follow-up samples of 5 out of 256 cyogenetically normal (CN-) AML who were FLT3 wild-type at presentation, but tested ITD+ at relapse or disease progression. Our study revealed that all patients carried a small ITD+ clone at diagnosis, which was undetectable by routine analysis (0.2–2% abundance). The dynamics of ITD+ clones from diagnosis to disease progression, assessed by UDS, reflected clonal evolution under treatment pressure. UDS appears as a valuable tool for FLT3 mutational screening and for the assessment of minimal residual disease (MRD) during follow-up, by detecting small ITD+ clones that may survive chemotherapy, evolve over time and definitely worsen the prognosis of CN-AML patients.

Karyotype evolution and acquisition of FLT3 or RAS pathway alterations drive progression of myelodysplastic syndrome to acute myeloid leukemia
We identified 38 patients (11 female, 27 male) who were analyzed by cytomorphology and cytogenetics both at diagnosis of MDS and later at progression to s-AML.

All 76 samples were analyzed by next-generation sequencing or polymerase chain reaction with a 33-gene panel targeting ASXL1, BCR, BRAF, CBL, DNMT3A, ETV6, EZH2, FLT3 (FLT3-ITD and FLT3-TDK), GATA1, GATA2, IDH1, IDH2, JAK2, KIT, KRAS, MLL-PTD, MPL, NPM1, NRAS, PHF6, RAD21, RUNX1, SETBP1, SF3B1, SMC1A, SMC3, SRSF2, STAG2, TET2, TP53, U2AF1, WT1, and ZRSR2.

In total, 15/20 patients (75%) who acquired new mutations in the evolution showed mutations in the signal transduction proteins (FLT3 or RAS pathway), indicating that these might be mutations driving s-AML transformation. FLT3 and NRAS mutations are thought to be important genetic events contributing to the pathogenesis of AML and the expected increase in the frequencies of mutations in s-AML cases was observed, confirming previously reported data.

Our data suggest that different underlying molecular mechanisms drive the progression from MDS to s-AML. On the one hand, karyotype evolution has an important impact on s-AML transformation. On the other hand, several mutations including those in ASXL1, ETV6, GATA2, IDH2, NRAS, RUNX1, and SRSF2 predispose to transformation to s-AML. However, mutations in signal
transduction genes (FLT3, KRAS, and NRAS) seem to drive the progression from MDS to s-AML more quickly if mutated at a specific time-point and should, therefore, be considered as prognostically informative during the disease course.

Acute Lymphoblastic Leukemia

Complex genetic heterogeneity influences prognosis in adult B-Cell Precursors Acute Lymphoblastic Leukemia negative for recurrent fusion genes

In order to better molecularly dissect this ALL subgroup, we performed an integrative molecular approach including gene candidate high-resolution screening and genome-wide profiling analyses.

We retrospectively analyzed 28 newly diagnosed BCR-ABL1-negative BCP-ALL subjects (19 males/9 females; median age 41.5 years; negative for known fusion genes) and 28 BCR-ABL1-positive BCP-ALL subjects as a comparison group. Overall, 76% of BCR-ABL1-negative subjects showed an abnormality of at least one of the analyzed known leukemia genes: 7 (25%) had one, 4 (14%) had two, 6 (21%) had three, and 6 (21%) had four or more alterations (WES data). In subjects showing no abnormalities, SNP arrays analysis revealed amplifications of chromosome 1q in 2/6 cases (33%). Deletions of CDKN2A/B were the most frequent (39%) and in 73%, they occurred together with other abnormalities, suggesting that multiple events are needed to induce the full leukemia phenotype. Other common CNA included: deletions of IKZF1 (25%), ETV6 (25%), PAX5 (14%), EBF1 (11%), PAR1 region (11%) and RB1 (7%). NGS showed mutations of JAK2 and CRLF2 in 7% (R683S/G) and 4% (F232C), respectively. No positivity for newly described fusion genes activating tyrosine kinase was confirmed. Importantly, subjects with no abnormalities showed better survival rates compared to those with one or more molecular alterations (p less than 0.01). The BCR-ABL1-positive subgroup shared the same CNA of BCR-ABL1-negative cases, such as deletions of IKZF1 (71%), CDKN2A/B (21%), PAX5 (14%), BTG1 (11%), EBF1 (11%), and ETV6 (4%), but they did not show mutations in the genes analyzed with targeted sequencing.

BCP-ALL lacking recurrent fusion genes is a highly heterogeneous and complex disease. Current diagnostic procedures need to be revised to improve risk assessment and to guide therapeutic decisions.

Clinical relevance of low burden BCR-ABL1 mutations detectable by amplicon deep sequencing (DS) in Philadelphia-positive (Ph+) Acute Lymphoblastic Leukemia (ALL) patients: the type of mutation matters

In Ph+ ALL patients treated with tyrosine kinase inhibitors (TKIs), the likelihood of acquiring TKI-insensitive mutations and the striking incidence of highly resistant T315I and compound mutants underscore the importance of BCR-ABL1 kinase domain (KD) sequence surveillance for timely and rational therapeutic reassessment.

We used an amplicon DS strategy of the BCR-ABL1 KD to assess the following issues: i) whether DS allows earlier detection of emerging TKI-insensitive mutations in patients undergoing BCR-ABL1 KD mutation screening for minimal residual disease (MRD) persistence; ii) whether TKI-insensitive low burden mutations can be identified in relapsed patients with negative conventional sequencing results; iii) whether TKI-insensitive low burden mutations are necessary and sufficient to predict for treatment failure in all cases.

This study was conducted in a total of 56 Ph+ ALL patients who received TKI-based therapies at our or collaborating institutions and were referred to our laboratory for MRD follow-up monitoring by RQ-PCR and for BCR-ABL1 KD mutation analysis in case of MRD positivity. MRD persistence in Ph+ ALL patients may hide emerging TKI-insensitive BCR-ABL1 KD mutations that DS may identify earlier than conventional sequencing - allowing a greater leeway before overt hematologic relapse occurs. Polyclonal resistance sustained by multiple TKI-insensitive low burden mutations may explain relapse in a proportion of cases with un-mutated BCR-ABL1 KD sequences as assessed by conventional sequencing. Moreover, the type of mutation matters: detection of low burden mutations insensitive to the ongoing TKI was always found to predict/correlate with treatment failure. Detection of low burden mutations with low/unknown IC50 might explain low level MRD but does not predict for an impending relapse.

MRD-triggered, BCR-ABL1 KD mutation screening by DS may be precious for earlier and more effective use of preemptive rescue therapies.

Essential Thrombocytemia
Screening of JAK2 V617F and MPL W515 K/L negative essential thrombocythaemia patients for mutations in SESN2, DNAJC17, ST13, TOP1MT, and NTRK1.

The most common mutations in Essential thrombocythemia (ET) are JAK2 V617F and MPL W515K/L, found in only about 60% of cases. In a recent study by Hou et al., single cells derived from a JAK2 V617F-negative ET patient were sequenced and eight other genes, which have not previously implicated in ET, were identified as possible candidate drivers. However, their recurrence rate in ET was not established. In our study we sequenced DNA from a series of 64 JAK2 V617F-negative and MPL W515K/L-negative ET cases for the reported mutations in SESN2, TOP1MT, ST13, DNAJC17, and NTRK1. None of these mutations were detected in our patients. However, we identified a novel acquired heterozygous mutation in TOP1MT (c.1400A>G p.N467S) by screening 102 ET patients. In silico analysis suggests that this mutation might affect the interaction of TOP1MT with the DNA molecule. In conclusion TOP1MT mutations may be recurrent in ET but at a low frequency.

Systemic Mastocytosis


Somatic mutations in the KIT receptor kinase (most frequently, D816V) can be detected in >90% of patients affected by Systemic Mastocytosis (SM) and are thought to play an important pathogenetic role. Indolent Systemic Mastocytosis (ISM) is the most common variant of SM, characterized by a very low MC burden and associated with very different clinical pictures. A highly sensitive diagnostic methods for D816V detection are required to assure an appropriate diagnosis and to reduce false-negative results. The recent development of "ultra-deep amplicon sequencing" (UDS) technologies has opened the way to a more accurate characterization of molecular aberrations with higher sensitivity of screening for known and unknown mutations.

Our aims were: i) to set-up and optimize a UDS-based mutation screening strategy of the KIT gene on the Roche GS Junior Instrument; ii) to test the sensitivity of our UDS assay to detect the D816V mutation; iii) to investigate the presence of additional KIT mutations in SM. We decided to take advantage of a next generation sequencing approach to perform an UDS KIT gene mutation analysis on 20 bone marrow (BM) samples from patients with ISM that were negative for the D816V mutation by Sanger Sequencing which has a sensitivity of 20%.

Two additional sequence variations in the c-KIT gene were detected in a large proportion of patients. These two variations included a 3bp in-frame deletion in exon 15 found in 11/20 patients and a 12bp in frame-deletion in exon 9 in all patients, with an abundance ranging from 83% to 97%.

Interestingly, our results showed the presence of the transmembrane domain M541L KIT-activating mutation in exon 10, with an abundance of 50%, in addition to D816V, in 2/20 ISM. This mutation is known to retain sensitivity to imatinib mesylate.

Chronic Lymphocytic Leukemia

Detailed analysis of therapy-driven clonal evolution of TP53 mutations in chronic lymphocytic leukemia.

In chronic lymphocytic leukemia (CLL), the worst prognosis is associated with TP53 defects with the affected patients being potentially directed to alternative treatment. Therapy administration was shown to drive the selection of new TP53 mutations in CLL. Using ultra-deep next-generation sequencing (NGS), we performed a detailed analysis of TP53 mutations' clonal evolution. We retrospectively analyzed samples that were assessed as TP53-wild-type (wt) by FASAY from 20 patients with a new TP53 mutation detected in relapse and 40 patients remaining TP53-wt in relapse. Minor TP53-mutated subclones were disclosed in 18/20 patients experiencing later mutation selection, while only one minor-clone mutation was observed in those patients remaining TP53-wt (n=40). We documented that (i) minor TP53 mutations may be present before therapy and may occur in any relapse; (ii) the majority of TP53-mutated minor clones expand to dominant clone under the selective pressure of chemotherapy, while persistence of minor-clone mutations is rare; (iii) multiple minor-clone TP53 mutations are common and may simultaneously expand. In conclusion, patients with minor-clone TP53 mutations carry a high risk of mutation selection by therapy. Deep sequencing can shift TP53 mutation identification to a period before therapy administration, which might be of
particular importance for clinical trials.

Recurrent mutations refine prognosis in chronic lymphocytic leukemia

Through the European Research Initiative on chronic lymphocytic leukemia (CLL) (ERIC), we screened 3490 patients with CLL for mutations within the NOTCH1 (n=3334), SF3B1 (n=2322), TP53 (n=2309), MYD88 (n=1080) and BIRC3 (n=919) genes, mainly at diagnosis (75%) and before treatment (>90%). BIRC3 mutations (2.5%) were associated with un-mutated IGHV genes (U-CLL), del(11q) and trisomy 12, whereas MYD88 mutations (2.2%) were exclusively found among M-CLL. NOTCH1, SF3B1 and TP53 exhibited variable frequencies and were mostly enriched within clinically aggressive cases. Interestingly, as the timespan between diagnosis and mutational screening increased, so too did the incidence of SF3B1 mutations; no such increase was observed for NOTCH1 mutations. Regarding the clinical impact, NOTCH1 mutations, SF3B1 mutations and TP53 aberrations (deletion/mutation, TP53ab) correlated with shorter time-to-first-treatment (P<0.0001) in 889 treatment-naive Binet stage A cases. In multivariate analysis (n=774), SF3B1 mutations and TP53ab along with del(11q) and U-CLL, but not NOTCH1 mutations, retained independent significance. Importantly, TP53ab and SF3B1 mutations had an adverse impact even in U-CLL. In conclusion, we support the clinical relevance of novel recurrent mutations in CLL, highlighting the adverse impact of SF3B1 and TP53 mutations, even independent of IGHV mutational status, thus underscoring the need for urgent standardization/harmonization of the detection methods.

Targeted next-generation sequencing in chronic lymphocytic leukemia: a high-throughput yet tailored approach will facilitate implementation in a clinical setting

Next-generation sequencing has revealed novel recurrent mutations in chronic lymphocytic leukemia, particularly in patients with aggressive disease. Here, we explored targeted re-sequencing as a novel strategy to assess the mutation status of genes with prognostic potential. To this end, we utilized HaloPlex targeted enrichment technology and designed a panel including nine genes: ATM, BIRC3, MYD88, NOTCH1, SF3B1 and TP53, which have been linked to the prognosis of chronic lymphocytic leukemia, and KLHL6, POT1 and XPO1, which are less characterized but were found to be recurrently mutated in various sequencing studies. A total of 188 chronic lymphocytic leukemia patients with poor prognostic features (un-mutated IGHV, n=137; IGHV3-21 subset #2, n=51) were sequenced on the HiSeq 2000 and data were analyzed using well-established bioinformatics tools. Using a conservative cutoff of 10% for the mutant allele, we found that 114/180 (63%) patients carried at least one mutation, with mutations in ATM, BIRC3, NOTCH1, SF3B1 and TP53 accounting for 149/177 (84%) of all mutations. We selected 155 mutations for Sanger validation (variant allele frequency, 10-99%) and 93% (144/155) of mutations were confirmed; notably, all 11 discordant variants had a variant allele frequency between 11-27%, hence at the detection limit of conventional Sanger sequencing. Technical precision was assessed by repeating the entire HaloPlex procedure for 63 patients. Concordance was found for 77/82 (94%) mutations. In summary, this study demonstrates that targeted next-generation sequencing is an accurate and reproducible technique potentially suitable for routine screening, eventually as a stand-alone test without the need for confirmation by Sanger sequencing. Therefore, we demonstrated that targeted NGS can be implemented in the clinical practice and we showed its applicability and reliability for detection of clinically relevant mutations.

Leukemia diagnostic panels

The natural consequence of the NGS-PTL high-throughput sequencing results was the set up and/or testing of leukemia panels that may help diagnostics, patients’ prognostication and stratification. In details, both a myeloid and a lymphoid panel were implemented by the consortium.

Myeloid malignancies

The consortium tested a novel NGS panel developed by Illumina, which covers mutational hotspots of 54 genes relevant to myeloid diseases (AML, CML, MDS, MPNs, CMML and JMML). It covers 15 full genes (exons only) and oncogenic hotspots in 39 additional genes, for a total of 568 amplicons of ~250 bp. This panel contains all the relevant genes highlighted in the WP4 and WP5 in myeloid leukemia and other recurrently mutated genes, according to literature. The optimization of the Illumina panel is ongoing. Moreover, additional custom panels are currently used by members of the NGS-PTL consortium (MLL) in order to define the genomic landscape of myeloid malignancies at diagnosis/relapse and select molecular biomarkers for follow up and
Diagnostic and prognostic Utility Of a 26-gene panel for Deep-Sequencing mutation analysis in Myeloid Malignancies

A comprehensive pan-myeloid panel to simultaneously target mutations in 26 genes allows a comprehensive analysis with the perspective to detect disease-defining mutations in the majority of patients.

We developed sensitive next-generation deep-sequencing (NGS) assays comprising in total 26 genes: ASXL1, BCOR, BRAF, CBL, DNMT3A, ETV6, EZH2, FLT3 (TKD), GATA1, GATA2, IDH1, IDH2, JAK2, KIT, KRAS, MPL, NPM1, NRAS, PHF6, RUNX1, SF3B1, SRSF2, TET2, TP53, U2AF1, and WT1. With the exception of RUNX1, which was sequenced on the 454 Life Sciences NGS platform (Branford, CT), all remaining genes were studied using a combination of a microdroplet-based assay (RainDance, Lexington, MA) and the MiSeq sequencing instrument (Illumina, San Diego, CA). The assay’s turn-around time was less than 6 days, loading up to eight patients per sequencing run. Thus far, 191 prospectively collected cases have been analyzed during routine operations. In all cases the assay was successfully performed. The major disease categories were as follows: MDS (n=76), suspected MDS (n=28), MDS/MPN (n=10), reactive bone marrow conditions (n=46), AML (n=8), CML (n=3), other conditions (n=20). A pan-myeloid screening assay using NGS allows addressing 26 relevant gene mutations in myeloid malignancies with diagnostic or prognostic impact. This approach is scalable and adoptable to accommodate the inclusion of novel gene targets according to the latest evidence from the literature. Importantly, given the broad spectrum of mutations in myeloid diseases covered by such a panel, mutations can be identified in the majority of patients and enable to support a more comprehensive classification in these complex diseases.

Lymphoid malignances

A 13-Gene Panel Targeted To Investigate CLL By Next-Generation Amplicon Deep-Sequencing Can Be Successfully Implemented In Routine Diagnostics

We developed a sensitive deep-sequencing assay adoptable to adjust gene targets and amplicons according to current state-of-the-art evidence regarding the published landscape of mutations in CLL. In total, 13 genes with relevance in CLL providing in part adverse molecular prognostic information were chosen: ATM, BIRC3, BRAF (V600), FBXW7, KLHL6, KRAS, NOTCH1 (PEST domain), NRAS, MYD88, POT1, SF3B1 (HEAT domain), TP53, and XPO1. Targets of interest comprised either complete coding gene regions or hotspots. The sequencing library was constructed starting off 2.2 μg genomic DNA per patient using a single-plex microdroplet-based assay (RainDance, Lexington, MA). Sequencing data was generated using the MiSeq instrument (Illumina, San Diego, CA) loading up to 10 patients per run. The total turn-around time of the assay was less than 5 days. In the cohort of 18 cases, a total of 71 mutation analyses had already been previously performed for eight of the 13 genes using either capillary Sanger sequencing or alternative amplicon deep-sequencing assays (454 LifeSciences or Illumina MiSeq). In detail, in these 8 genes these 71 assays detected 56 known polymorphisms or mutations in ATM (n=8), BIRC3 (n=6), FBXW7 (n=4), MYD88 (n=4), NOTCH1 (n=10), SF3B1 (n=5), TP53 (n=14), and XPO1 (n=4) and 28 analyses revealed a wild-type status. When comparing these results with data obtained using the 13-gene NGS panel, in all 84/84 (100%) parallel assessments concordant results were obtained underlining the robustness of this assay. Importantly, a number of patients (14/18) was detected to harbor mutations in genes reported to be associated with decreased overall survival, both in high-risk (e.g. TP53, BIRC3) and intermediate-risk (NOTCH1, SF3B1) categories according to Rossi et al., 2013. As such, detecting these adverse somatic alterations may influence the course of therapy for these patients underlining the utility of such a screening panel.

Inclusion of relevant genomic alterations into clinical practice

One of the main goals of genomic analyses is represented by their applicability in everyday clinical practice as markers of treatment response. The analysis of a great number of leukemia cases, as performed by NGS-PTL consortium, favored the introduction of specific genomic analyses as inclusion criteria of clinical trials and during follow up in the regular clinical practice.
Complex karyotype, older age, and reduced first-line dose intensity determine poor survival in core binding factor acute myeloid leukemia patients with long-term follow-up. Approximately 40% of patients affected by core binding factor (CBF) acute myeloid leukemia (AML) ultimately die from the disease. Few prognostic markers have been identified. We reviewed 192 patients with CBF AML, treated with curative intent (age, 15-79 years) in 11 Italian institutions. Overall, 10-year overall survival (OS), disease-free survival (DFS), and event-free survival were 63.9%, 54.8%, and 49.9%, respectively; patients with the t(8;21) and inv(16) chromosomal rearrangements exhibited significant differences at diagnosis. Despite similar high complete remission (CR) rate, patients with inv(16) experienced superior DFS and a high chance of achieving a second CR, often leading to prolonged OS also after relapse. We found that a complex karyotype (i.e. ≥4 cytogenetic anomalies) affected survival, even if only in univariate analysis; the KIT D816 mutation predicted worse prognosis, but only in patients with the t(8;21) rearrangement, whereas FLT3 mutations had no prognostic impact. We then observed increasingly better survival with more intense first-line therapy, in some high-risk patients including autologous or allogeneic hematopoietic stem cell transplantation. In multivariate analysis, age, severe thrombocytopenia, elevated lactate dehydrogenase levels, and failure to achieve CR after induction independently predicted longer OS, whereas complex karyotype predicted shorter OS only in univariate analysis. The achievement of minimal residual disease negativity predicted better OS and DFS. Long-term survival was observed also in a minority of elderly patients who received intensive consolidation. All considered, we identified among CBF AML patients a subgroup with poorer prognosis that might benefit from more intense first-line treatment.

Analysis of phenotype and outcome in essential thrombocythemia with CALR or JAK2 mutations

The JAK2 V617F mutation, the thrombopoietin receptor MPL W515K/L mutation and the Calreticulin (CALR) mutations are mutually exclusive in essential thrombocythemia and support a novel molecular categorization of essential thrombocythemia. CALR mutations account for approximately 30% of essential thrombocythemia cases. In a retrospective study, we have examined the frequency of MPL and CALR mutations in JAK2 V617F negative essential thrombocythemia (n=103). In addition, we compared the clinical phenotype and outcome of CALR mutant essential thrombocythemia with a cohort of JAK2 V617F positive essential thrombocythemia (n=57). CALR positive cases represented 63.7% of double negative essential thrombocythemia, and most carried CALR Type 1 or Type 2 indels. However, we also identified one patient, who was positive for both the JAK2 V617F and the CALR mutations. This study revealed that CALR mutant essential thrombocythemia is associated with younger age, higher platelet counts, lower erythrocyte counts, lower leukocyte counts, hemoglobin, hematocrit, and increased risk of progression to myelofibrosis in comparison with JAK2 V617F positive essential thrombocythemia. Analysis of the CALR mutant group according to indel type showed that CALR Type 1 deletion is strongly associated with male gender. CALR mutant patients had a better overall survival than JAK2 V617F positive patients, in particular patients of age 60 years or younger. In conclusion, this study on a Belgian cohort supports and extends the growing body of evidence that CALR mutant is phenotypically distinct from JAK2 V617F positive essential thrombocythemia, with regard to its clinical and hematological presentation as well as the overall survival.

PKC412 (Midostaurin) is safe and highly effective in systemic mastocytosis patients: Follow up of a single-center Italian compassionate use

Treatment of SM usually focuses on symptom relief by histamine receptor antagonists and other supportive therapy. However, in aggressive and leukemic variants, cyto reducective and targeted drugs must be applied. Thus, from March 2011, 9 (M/F = 3/9) patients with ASM have been treated with PKC412, administered orally, at the dosage of 100 mg twice daily, continuously. The median age was 60 years (range 39-75); the median time from diagnosis was 6 months (range 2-53). Median serum tryptase level was 100 mcg/L (range 19.3-1160). C-kit mutation D816V was present in 8 out of 9 patients. Cytogenetic analysis was normal in all the patients. According to European Criteria, a Major response was observed in one patient, and a partial response in 6 patients. Overall, the drug was well tolerated, and no serious adverse events were observed. All the patients obtained a quick improvement of clinical symptoms, in terms of weight gain, bowel function and skeletal pain. At the bone marrow evaluation, the persistence of the D816V c-kit mutation was observed, despite a significant decrease of mast cell marrow involvement. Conclusions: In a small cohort of ASM patients, the prolonged therapy with PKC412 is safe and effective, mainly on symptoms improvement and
haematological profile. Nevertheless, the persistence of the D816V c-kit mutation suggests that many other oncogenic factors may be responsible for the pathogenesis of the disease.

Extremely high rate of complete hematological response of elderly Ph+ acute lymphoblastic leukemia (ALL) patients by innovative sequential use of Nilotinib and Imatinib. A GIMEMA Protocol LAL 1408

We have explored if the administration of two TKIs, Nilotinib (NIL) and Imatinib (IM) can improve the results without increasing the toxicity in the elderly Ph+ Acute Lymphoblastic Leukemia (ALL) patients. We investigate the type and number of BCR-ABL kinase domain mutations developing during and after the study. 39 patients have been enrolled in 15 Italian hematologic Centers (median age 66 years, range 28-84). Among these, 8 patients were unfit for standard chemotherapy or SCT (median age 50 years, range 28-59). 27 patients were p190, 5 were p210 and 7 were p190/p210. After 6 weeks of treatment, 36 patients were evaluable for response: 34 were in CHR (94%) and 2 in PHR (6%). 23 patients have already completed the study core (24 weeks), 87% were in CHR and 17 are currently continuing therapy in the protocol extension phase. Thus, the OS at 1 year is 79%, and 64% at 2 years. Overall, 1 patient was primarily resistant and 13 patients have relapsed, with a median time to relapse of 7.6 months (range 0.8-16.1 months), for a DFS of 51.3% at 12 months. Conclusions: In this small cohort of Ph+ ALL elderly/unfit patients, the rates of relapse and progression were not likely to be different from the rates observed with Imatinib alone.

In addition to the results reported above, the table shows the genomic biomarkers tested in clinical trials (both spontaneous and company-sponsored) performed by consortium partners in the last year or that are going to open the enrollment phase in the next months (see attached file).

Conclusions

The development and introduction of an increasing number of targeted therapies is strictly related to the technological advancement in the diagnostic field and the expansion of the molecular and biological knowledge. Indeed, genomic-based mechanisms of drug resistance through clonal selection and clonal evolution represent a major challenge of the novel therapies. On the other hand, genetic alterations help select potential sensitivity to be exploited in novel clinical trials in order to define patients’ subgroups that are expected to benefit from inhibition of disease specific pathways. Indeed, the identification of specific biomarkers involved in malignant transformation and/or drug resistance paves the way for further validation across European patient cohorts and will facilitate an early and effective patient access to innovative and targeted treatment approaches.

A better understanding of the molecular complexity of leukemia, of the disease natural history and of its real-life consequences, is the key tool to guide decision-making at any level, including clinicians wishing to provide tailored therapeutic options to patients, pharmaceutical industries investing in drug design, regulatory agencies evaluating drug development process and availability and other stakeholders in the healthcare system. Therefore, the collection of a huge number of clinically and molecularly characterized leukemia cases is a valuable source for the European scientific and clinical community, facilitating future data interpretation. Moreover, the scientific and technological achievements of the NGS-PTL consortium, including protocols, pipelines, data, knowledge and storage solutions provide the bases to speed up effective changes in everyday clinical practice, starting from a discovery phase at disease diagnosis/relapse and moving deeply to targeted analyses at follow up, which revealed a great clinical utility. The obtained results will help prioritize therapeutic interventions and foster the development of innovative therapeutic approaches, which are needed for most leukemia cases and in particular for patients characterized by high-risk profiles and intrinsic resistance to conventional therapies.

References


Potential Impact:
Hematological diseases account for approximately 9.5% of the new diagnosed cancers every year, with an overall incidence, including both acute and chronic forms, that is slightly higher in subjects of European ancestry in comparison to those belonging to other ethnic groups, and reaches a value of 9.6 cases per 100,000 individuals. Similarly to several solid tumours, the incidence of acute leukemia shows an exponential rise after the age of 40, being thus highly impacting on elderly populations. Considering the impressive and continuous increase in life expectancy that is recorded worldwide and in particular in Europe, where the proportion of individuals aged over 65 years old is becoming greater and greater and is expected to nearly double within 2030, hematological diseases promise to become a substantial burden for the healthcare systems of present-day industrialized societies, especially of the European one. Accordingly, the costs and the severe impact that the currently adopted anti-leukemia therapies have on patients’ quality of life and well-being, even when the disease is successfully fought, will became a primary challenge for the European health care systems.

Results obtained within the framework of the NGS-PTL project are thus expected to substantially enhance the health and quality of life of individuals affected by several leukemia subtypes and, at large, of the European population as a whole. In fact, the development of new tools for enabling early-diagnosis and prognosis of several leukemias, as well as for evaluating the effectiveness of specifically tailored therapeutic interventions, was achieved by project activities through innovative massive parallel sequencing approaches and is expected to improve the translation of personalized medicine into clinical practice, thus contributing to a greater sustainability and efficiency of the healthcare system, due to the fact that very expensive, but in some cases useless or even harmful, therapies could be avoided. In particular, the project succeeded in identifying and fine-tuning an experimental framework that for the first time enables the depiction of a comprehensive disease fingerprint for each patient, thus laying the foundation for a concrete personalized medicine approaches. Moreover, the project succeeded in organizing the first European clinical and technological platform for the study, diagnostic and cure of several leukemia subtypes by means of innovative sequencing technologies and analytical methods, thus enabling to systematically draw the as exhaustive as possible picture of each patient’s genome complexity in the attempt to provide more effective guidance to routine leukaemia diagnosis, prognosis and treatments.
As a matter of fact, results obtained within the project relied on the detection of leukemia individuals’ exomic and transcriptomic signatures that led to the identification of novel diagnostic and prognostic biomarkers which are expected to improve the selection of tailored therapeutic approaches. As a consequence, both anti-leukemia therapeutic success and sustainability and efficiency of the healthcare system are expected to take advantage from the results achieved with the project.
In addition to this issue, the network of international collaborations established and consolidated during the project has enabled to create, and will also further develop, a web-based platform including exomic, transcriptomic and phenotypic/clinical information associated to each examined patient, thus facilitating the sharing and interpretation of the generated “omics” data and providing a precious tool also for the whole scientific community devoted to the study of haematological malignancies.
The performed analyses of such large cohorts of leukemia patients, coupled with the correlation of obtained biological and...
omics data with clinical ones, has improved not only basic research on the main factors influencing the ethiogenesis and
development of a variety of haematological diseases, but also the definition of relevant biomarkers involved in malignant
transformation and, especially, in drug resistance. This was achieved by the depiction of exomic and transcriptomic profiles of
each patients and by subsequent validation of the most promising signatures and/or biomarkers on independent disease
samples. According to this approach, potential application of the generated knowledge to the molecular diagnostic field will
directly impact future clinical trial structures and patient recruitments towards rigorous interventional studies addressing very
specific questions that pay the way to precision personalized medicine. The concrete implementation of such a theoretical
framework will mainly result in a twofold benefit:

- Very expensive and in some cases useless or even harmful therapies will be avoided.

- New biological drugs will be specifically directed only to patients for whom they are safe and useful.

According to this view, scientific results obtained by means of the project activities are expected to set the stage for further
research programs aimed at implementing concrete personalized medicine approaches, as envisage by EC and in particular
devoted to:

- Improve the development and design of innovative products and services for diagnosis, prognosis and cure of
haematological diseases, also by involving industrial players in the process of selection and implementation of the most
relevant and cost-effective usable biomarkers;

- Establish more sustainable health and care systems thanks to the adoption of the most suited and really effective treatments
for each patient, thus avoiding potential short and long-term side effects and consequently reducing subsequent patients’
hospitalizations and related high costs;

- Ensure both benefits for leukemias patients and cost-savings from not administering ineffective/toxic drugs thanks to the
possibility of obtaining each individual disease fingerprint by means of rapid and cost-effective screening based on the most
innovative massive parallel sequencing technologies;

- Represent an overall ground-breaking approach to fight human cancers by provide new models useful for research also on
other neoplastic diseases in addition to leukemias, thus promoting broader innovation in targeted therapies and providing a
greater treatment effect in patients by triggering massive parallel sequencing-based discovery of therapeutic targets and
patients’ stratification procedures in several malignancies.

As a matter of fact, personalised medicine represents one of the top priorities for the EU, being substantially advanced by
Horizon 2020. Although the NGS-PTL project was launched before the this research programme, the scientific milestones
achieved within it are fully in line the objectives scheduled with Horizon 2020 for personalized medicine as they have
contributed to:

- Improve the benefits for patients thanks to early diagnosis and identification of the most appropriate and tailored therapeutic
approaches;

- Improve the sustainability of health-care systems due to cost-savings associated to not administering ineffective/toxic drugs;

- Promote innovation in drugs that are targeted and will provide a greater treatment effect in patients who respond.

The project results demonstrate that technological and methodological tools are already available to actually drive the
adoption of the most suited and really effective treatments for each leukemia patient, thus opening the way for the
overcoming of potential short and long-term side effects and for reducing subsequent patients’ hospitalizations and related high costs.

In fact, the findings described by the project, as well as the established experimental and methodological frameworks devoted to enable accurate patients stratification at the time of diagnosis will directly impact on future management of subject affected by haematological malignancies by assisting clinicians in making the best decision with respect to the disease efficient and sustainable treatment.

As mentioned before, resulting benefits will be not strictly restricted to patients with hematological diseases, as these malignancies have traditionally represented excellent models also for research on different neoplastic disorders. Results obtained by the NGS-PTL project have thus the potential to massively promote and strengthen research and therapeutic efforts on several other cancers, by developing a methodological framework devoted to the application of massive parallel sequencing approaches to the discovery of therapeutic targets and patients’ stratification procedures in different malignancies and thus potentially representing a ground-breaking approach in fighting human cancers in general.

As expected, the project generated both scientific knowledge and concrete products (i.e. leukemia diagnostic panels) able to enhance the health and quality of life of leukemia patients, by laying the foundation for the development of new tools for early-diagnosis and prognosis of several leukemia subtypes, and putting a step forward to the translation of personalized medicine into routine clinical practice. The project indeed strongly contributed to the improvement of scientific management of huge amounts of exomic and transcriptomic data from which extract specific information useful to routine clinical practice and to provide an incisive guidance to therapeutic interventions.

At the same time, the NGS-PTL project also boosted industry research, in particular as concerns the development and application of specific massive parallel sequencing-based approaches to oncological research. In the mid and long term, this opened the way for completely new business opportunities for the SMEs involved in the project thanks to the highly specific expertise that they acquired and/or consolidated during the project. In fact, they had the chance to become some of the leading groups in the European market of massive parallel sequencing applications to clinically-oriented genomic analyses, by having developed and validated during the project a series of effective protocols and tools suitable for supporting clinical and academic research groups involved in the study and care of leukemias and other cancers. Moreover, their acquired facilities and expertise have now the potential to be easily adapted to the requirements of broad oncological research as a whole, thus ensuring them a pivotal and profitable role in the upcoming and momentous process of translation of omics-based personalized medicine into routine clinical practice.

In conclusion, both the scientific and technological results achieved within the NGS-PTL project contributed to consolidate the paradigm shift from standard clinical approaches of evidence-based medicine to pioneering personalised medicine, in which identification of specific biomarkers and therapeutic targets actually guides treatment decisions towards innovative targeted therapies.

According to this, the project has had, and will continue to have in the next years, a strategic impact on the European society as a whole. In fact, the innovative products that it has developed, coupled with the methodological processes that it has set up, will directly contribute to several EU’s societal objectives, especially those concerning the citizens’ quality of life and health, as well as the sustainability of the healthcare services.

The main impact of the project results on these specific issues is outlined as follows:

- As regards the citizens’ quality of life, the project results lay the foundation for identifying the as suitable as possible therapeutic treatments that could be disposed only to patients for whom they are safe and useful, thus avoiding potential harmful side-effects and consequently reducing the severe impact that the currently adopted anti-leukemia therapies have on patients’ quality of life and well-being.

- As regards the citizens’ health, the project produced a relevant amount of scientific knowledge, as well as of targeted diagnostic panels, that will enable significant increase in the proportion of leukemia patients who will be earlier diagnosed, thus giving them higher chance of recovery and also preventing the use of too aggressive therapies that would be necessary for late-stage treatments and that may result in increased side effects.
As regards the sustainability of the healthcare services, accurate patients stratification enabled by the methods developed within the project is expected to reduce the treatment cost per patient since tailored approaches will ensure that very expensive, but in some cases useless or even harmful, therapies can be avoided and that subsequent patients’ hospitalizations due to therapies side effects will be reduced.

Main dissemination activities and exploitation of results

During the entire funding period the NGS-PTL partners have paid great attention to activities aimed at communicating and transfer the knowledge generated within the project to the scientific, commercial, and policy communities, as well as to the general audience.

For this purpose, an NGS-PTL webpage was firstly created to both provide a private access area to establish a safe platform for data and information exchange between the members of the consortium and at the same time to provide an open-access interface aimed at interactively communicating selected results outside the consortium, as well as at providing an overview of the consortium activity. For instance, this webpage progressively collected the main project information on publications and communications (e.g. papers, conferences, and exhibitions), being up-dated on a regular basis. Moreover, it summarized other information relevant to the project activities, allowing the public to get detailed information on the consortium partners and related projects. Information on the webpage will continue to be updated also after the completion of project activities to comprehensively summarize the final achievements of the consortium.

Along with this initiative, a substantial number of peer reviewed articles was published in leading journals of the field, as well as many presentations of the results obtained within the project were done at national and international conferences, local meeting and lectures. All of these dissemination activities contributed to largely increase the visibility of the NGS-PTL consortium.

In details, the academic partners of the consortium have published onto scientific journals novel insights into the development and progression of several leukemia subtypes. In details, the consortium as a whole has already produced 42 publications on internationally renowned journals. This unprecedented publication activity is far beyond what one would have expected within the funding period as many projects were just recently finished. In accordance, several additional manuscripts are currently in preparation so that a substantial number of manuscripts referring to the NGS-PTL findings will be published in the next few months.

The project partners have also presented results obtained by the consortium at several international events. In particular, members of the NGS-PTL project presented project-related findings in the context of the major worldwide cancer meeting focused on the study of hematological malignancies, such as the annual meetings of the American Society of Hematology (ASH), those of the European Association of Hematology (EHA), those of the American Association for Cancer Research (AACR), those of the American Society of Clinical Oncology (ASCO), as well as the European Society of Medical Oncology (ESMO). Moreover, the NGS-PTL consortium launched an “outreach”-program in a way that also the Arabian, South American or Asian haematology societies were informed of the NGS-PTL efforts during smaller international meetings. This also led to the opportunity of initiating novel research initiatives at the respective countries. In total, the members of the consortium presented the activities of the NGS-PTL network in more than 100 international scientific meetings.

Moreover, construction of a data warehouse including all clinical information and genomic data collected during the project will allow the research community to access the data generated within the project activities, thus representing a long-lasting invaluable resource to the whole scientific community. This will also provide a means to elaborate additional joint projects with industrial partners to further exploit the information already stored within the NGS-PTL repository.

In addition to the described traditional dissemination activities and to the implications arose from the creation of a publicly accessible data warehouse, several satellite projects have also originated from the NGS-PTL consortium, resulting in joint grant proposals and in the formation of joint initiatives, as well as in strong interaction with industry. This will further improve the actual implementation of the optimized massive parallel sequencing-based approaches to daily clinical routine.

As regards the exploitations of the obtained results, the description of the NGS-PTL schedule on how to use and further disseminate the findings of the project was summarized in the plan for the use and dissemination of foreground (PUDF). This
document was drafted by the appointed exploitation manager within the framework of the activities carried out by the Dissemination and Exploitation Task Force. While the first project findings resulted in patent applications, building up a data warehouse, which poses an invaluable asset to the research community, represents a further milestone achievement that will be further exploited within collaborative efforts in the future. In addition to this, the knowledge gathered within the network was further disseminated within a white paper on “dos and don’ts” in NGS data analysis. Furthermore, the NGS-PTL expertise was also exploited with regard to the ongoing development of diagnostic tests in collaboration with industrial partners (e.g. Illumina) and first molecular markers have been also entering clinical trials.

A full description of the exploitable foreground is reported as follows:

(I) White paper on “dos and don’ts” in NGS data analysis for molecular diagnosis

Based on the knowledge accumulated within the network, members of the NGS-PTL consortium have wrote a white paper to provide a guide that informs researchers interested on massive parallel sequencing concisely about the complex issues of this technology and, especially, on its applications for molecular diagnostics. This document presents the “NGS-PTL philosophy” on the matter, which is meant to help readers:

(i) to understand potential problems and pitfalls in the analysis of data generated by means of massive parallel sequencing experiments;
(ii) to solve specific technical problems related to the application of massive parallel sequencing approaches to the study of haematological malignancies;
(iii) to make decisions with regard to data interpretation.

Accordingly, this white paper recommends new and improved solutions to the nagging problem of how to best analyse complex data sets generated by means of massive parallel sequencing experiments and it will be best used to build mind share, inform and convince stakeholders of the effectiveness of these innovative technology if properly used. This in term might help to further negotiate reimbursement for molecular diagnostics based on the described experimental approaches, as in many countries so far massive parallel sequencing-based diagnostics is not yet funded by the public healthcare systems.

(II) Service/app for bioinformatics analysis, data interpretation and leukemia consulting

The NGS-PTL consortium will exploit the knowledge and expertise acquired and consolidated during the project activities to develop services and softwares to help haematological scientists with the bioinformatics analyses of data generated from massive parallel sequencing experiments, with data interpretation, as well as to support consulting activities to physicians and patients.

Specifically, good practice documents were created and made available for the scientific community and pharma industry. Bioinformatics pipelines optimized for the study of haematological malignancies were also made available on demand to laboratories in order to disseminate the knowledge achieved during the project and to propose a standardized and reproducible robust computational approach to detection of disease-involved variants and fusion transcripts. Mining of meaningful insights about the detected variants is enabled by scripts of free code developed to be portable (e.g. R environment for statistical computing) and relying on machine learning techniques for univariate and/or multivariate analyses, Bayesian networks and stochastic modeling.

Several documents are available to scientific community and pharma industries on demand. All data collected are stored in a data warehouse as described below and available for retrospective studies. Bioinformatics knowledge is available also as a service for clinician and biologists who need well-established computational pipelines for the analysis of whole exome and whole transcriptome sequencing data in the field of hematological malignancies research, as well as in diagnostics.

(III) NGS-PTL data warehouse – NGS data repository for explorative studies

During the funding period, the project consortium has successfully built up a web-based data warehouse that will provide an invaluable asset for the research community and for novel collaborative interactions with industrial partners.

In fact, similarly to the NIH data base dbGaP, the database of Genotypes and Phenotypes, the NGS-PTL data warehouse was developed to archive and distribute the results of studies that have investigated the interaction between genotype and
phenotype in the context of several different haematological malignancies.

Researchers and industrial collaboration partners interested in using the NGS-PTL data will simply have to fill in a request to the data management committee consisting of the NGS-PTL working group leaders, in which they state the reason, extent and purpose of the requested data analysis. Following a favourable vote by the NGS-PTL data management committee, an agreement will be signed between the consortium and the applicant institution and the respective collaborative partner will get access to the “non-private” data contained in the data warehouse. Access to the system/data warehouse, that is maintained by the consortium partner SINAPTICA, will be established through a link on the NGS-PTL webpage (see also https://www.sinaptica.it/clinicaltrials_client/?database=clinicaltrials_NGSPTL).

Information from the NGS-PTL repository can then be used to guide industry partners with regard to genes to be included in novel diagnostic panels.

In accordance to this policy, the NGS-PTL consortium has established a successful collaboration with Illumina and several project members have been already involved in inter-laboratory comparison studies for massive parallel sequencing-based targeted re-sequencing panels for the study of AML. In an on-going collaborative effort, the NGS-PTL network was also involved in further refining a diagnostic and prognostic AML panel and data generated within the consortium has been already demonstrated to represent resources helpful for determining the significance/relevance of the markers implemented within the panel.

(IV) Leukemia Exome SNVs Atlas

Variants detected in the exomes of patients sequenced within the NGS-PTL activities were grouped and organized such as a catalogue, which is going to be available for the scientific community and pharma industries. The database contains leukemia-associated variants derived from the sequenced leukemia patients, coupled with related clinical information, thus constituting a completely novel map of exome variants. This effort was focused to obtain a comprehensive catalogue of haematological malignancies-specific variants that could be useful to better characterize diseases and help to find novel diagnostic and prognostic markers.

This amount of knowledge is manually curated and all information are made available by web-based services. Moreover, a graphical interface (GUI) is under development to enable a user friendly browsing of the database content. Several statistical tools are also part of the GUI to provide basic information about variants, such as allele frequencies, association tests, regression and survival analyses.

The database will be updated as additional leukemia patients will be sequenced by NGS-PTL partners also after the completion of the project.

(V) Incorporation of novel biomarkers into clinical trials

During the funding period, the NGS-PTL consortium has also succeeded in starting novel clinical trials that include biomarkers not yet belonging to the standard of care. These markers have been shown either to represent targets for novel treatment approaches, such as the KIT gene in an AML treatment trial incorporating a tyrosine kinase inhibitor targeting its mutant version, or to provide a very good surrogate for treatment response, such as the MLL rearrangements that are linked to CDK4/6 deregulation, which in turn can be targeted by palbociclib.

Additional findings arose from the NGS-PTL activities (stemming also from on-going analyses) will provide further insights into novel biomarkers that can be investigated within upcoming prospective clinical trials that are currently being designed. This will advance the field of precision medicine as novel biomarker driven treatment approaches are more likely to provide a clinical benefit to the patients.

Ultimately, the findings of the NGS-PTL project as regards the identification of novel prognostic and predictive biomarkers will be also implemented in refined classification guidelines. For instance, the European and International Leukemia Classification Guidelines, such as the ELN or WHO classifications, are going to be revised in the near future and findings emerged from the NGS-PTL associated study groups are expected to impact this revised versions (e.g. anticipated inclusion of TP53 mutations in the AML classification).

Furthermore, these biomarkers might lay the foundation for the development of robust tests that can for instance predict response to novel targeted treatment strategies, such as in the case of development of routine KIT testing in all AML.
In accordance, the respective “novel biomarker tests” will be also of economic benefit as more effective guidance of expensive novel treatment approaches can significantly reduce health care costs. This, in turn, will provide additional resources necessary to treat more patients with targeted treatment strategies (e.g. 11 out of 12 recently approved cancer drugs cause an average cost of 100,000 US$/ treatment per year). Furthermore, this can significantly reduce health care costs, thereby providing additional resources to further improve supportive and quality of life measures. Finally, from a patient’s perspective, this approach will also minimize the number of “ineffective” treatment strategies, which will improve patient quality of life and might also translate it in improved patient outcome. Accordingly, these novel biomarkers and targeted re-sequencing approaches need to be further tested within clinical trials (in collaboration with industrial partners, see for instance the analysis of PIM gene mutations in a trial testing a PIM kinase inhibitor) as outlined above. Finally, the expertise of the consortium will be further exploited by designing novel plans for “basket studies”, i.e. treating leukemia/cancer patients based on underlying aberrations rather than “classical leukemia subgroups”. This will allow testing the value of novel biomarkers and treatment approaches in larger patient cohorts, thereby increasing the chance to identify favourable effects.

(VI) Development of additional NGS-based prediction tools
In accordance to the exploitable foreground mentioned at point V, many companies devoted to the development of massive parallel sequencing approaches recently started to focus their portfolio also on cancer in order to develop “commercial” cancer panels for improved diagnosis and risk prediction. For instance, the Illumina “TruSight myeloid panel”, has been developed in tight collaboration with the NGS-PTL partners. In addition to these predictive and prognostic tests, a large potential was ascribable also to the development of additional tools desperately needed to further the use of massive parallel sequencing technologies in the clinic practice. These include novel interactive applications that will also “lay” interpretation of sequencing data, that is by providing a tool that will allow a general physician, who are participating in the care taking of hematological patients, to also understand a molecular report and its direct impact on the patient management. This will allow him to better communicate to his patient and the tool itself might be also developed in a way that the patients can also further explore their individual tumour associated aberrations in a meaningful way.

In addition, the consortium has also further explored the generated data with regard to their value for the design of individualized minimal residual disease (MRD) monitoring tests. These would address the unmet medical need to have better disease monitoring tools, especially with regard to the better guidance of targeted therapeutic approaches, thereby making them more successful. Moreover, the development of a collection of information regarding clinical trials and target therapies available at different institutes and hospitals in Europe will be pursued (“Therapies Around Me”). The related database will be available for both physician and patients who are seeking for therapies and drugs which target specific leukemia biomarkers.

List of Websites:

http://www.ngs-ptl.com/default.aspx

Project Coordinator: Prof. Giovanni Martinelli
mail address: giovanni.martinelli2@unibo.it
Phone +39 051 390413 - Fax +39 051 398973

<table>
<thead>
<tr>
<th>Related information</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Result In Brief</strong></td>
<td>Personalised treatment for leukaemia</td>
</tr>
<tr>
<td><strong>Documents and Publications</strong></td>
<td>final1-clinical-trials.pdf</td>
</tr>
</tbody>
</table>