Collection of Executive Summary
relative to the Deliverables of the Second Report Period

D1.2 – Second List of recruited β-thal patients with characterized genotype/phenotype

Executive Summary
Deliverable D1.2 is an update of D1.1 and it related to Tasks 1.1 and 1.2. D1.2 reports the second THALAMOSS list of recruited patients for future activities employing blood sampling, culturing of erythroid cells and isolation of genomic DNA, RNA and protein. D1.2 incorporates data already delivered under D1.1 and contains the list of patients who are participating to the THALAMOSS. For post of the patients the genotype/phenotype was already known or has been determined under the THALAMOSS project. This list now includes 127 entries contributed by UNIFE, 356 entries by CING, 22 entries by CU, 100 entries by UNICA, 5 entries by KC and 334 entries by LGHA (total 944 entries). This list represent the number of β-thalassemia patients participating to the THALAMOSS activity. The analysis of the most frequent genotypes and phenotypes allows us to conclude that homozygous patients are 135 $\beta^0$39/$\beta^0$39, 210 β^+IVSI-110/β^+IVSI-110 and 21 β^+IVSI-6/β^+IVSI-6. Double heterozygous patients are 58 $\beta^0$39/β^+IVSI-110, 95 β^+IVSI-6/ β^+IVSI-110 and 36 β^0IVSI-1/β^+IVSI-110. In the list are present 36 homozygous sickle-cell anemia (SCA, HbS/HbS) patients. Finally, the HbS genotype was associated in 29 cases with $\beta^0$39, in 61 cases with β^+IVSI-110, in 13 cases with β^+IVSI-6 and in 17 cases with β^0IVSI-1.

D1.7 – Centralized β-thal cellular THALAMOSS BioBank

Executive Summary
In D1.7 we describe (a) updates on the generation of a β-thalassemia patients cellular biobank (Thal-Biobank), (b) its characterization in respect to maintenance of the phenotype (% of production of fetal hemoglobin), (c) its validation in terms of subculturing of frozen cryopreserved samples in different laboratories and induction of fetal hemoglobin using Hydroxyurea as fetal hemoglobin inducer. Hydroxyurea was chosen among the different available fetal hemoglobin (HbF) inducers, since it is already used in experimental therapy of patients affected by β-thalassemia and sickle-cell anemia. At present, the BioBank is constituted by ErPCs from 135 patients, and more than 688 cryovials.

D2.2 – SNP analysis completed of at least 100 genomic samples

Executive Summary
Genomic DNA samples from more than 300 β-thalassemia patients have been used for SNPs analysis identification of β-globin gene mutations, polymorphisms of the β-globin genes, the XmnI polymorphism of the promoter of the fetal γ-globin gene, and polymorphisms of the BCL11A and HBS1L-MYB loci, which are disease modifiers through their role in fetal γ–globin expression. A novel polymorphism was also found within the 5'UT sequence (+25) of the Aγ-globin gene. This Aγ(+25 G->A) polymorphism is associated with the Gγ-globin-XmnI polymorphism and both are linked with β039-globin

genes. The *KLF1* gene was sequenced in 29 selected β-thalassemia patients, including the promoter up to -200 from the transcription start site, the three exons and the two introns. Genomic variants were observed. In addition to known missense mutations in KLF1 that are believed to be of no functional consequence, a number of promoter and intronic variants of which the significance is not clear at the moment were found; detailed analysis in combination with the patient data in the THALAMOSS database may reveal a correlation of these variants with patient phenotypes.

D2.3 – RNA collection completed of at least 50 ErPC for WP2

ExecutiveSummary
92 blood samples were collected from patients with β-thalassemia intermedia (TI) and major (TM) and healthy controls, and, after RNA extraction, the samples were inserted in the THALAMOSS RNA data-bank. Collection of cells was performed at day 4 or day 8 of erythroid differentiation. To validate whether the RNA collected from β-thalassemia patients are suitable for transcriptomics profiling, RNA was analyzed with quantitative RTPCR for γ-globin and BCL11A-XL. To further verify in a genome wide level whether the RNA samples are suitable for transcriptomics profiling, microarray analysis was performed using RNAs from subjects expressing low or high levels of fetal hemoglobin (HbF). The data obtained clearly indicate that when low-HbF and high-HbF subjects are compared, a sharply different transcriptomic profile was obtained, allowing easy ranking and identification of genes which are down- or up-regulated in ErPCs expressing high levels of HbF. We conclude that the RNA samples collected and inserted in THALAMOSS RNA biobank are suitable for RNA-seq analysis.

D2.4 – RNA-seq completed of at least 50 RNA samples

Executive Summary
51 RNA samples collected from patients with β-thalassemia (intermedia, TI and major, TM) and inserted in the Thalamoss RNA data-bank were processed for transcriptomics analysis. The data obtained clearly indicate differences in transcriptomic profiles obtained in thalassemic patients versus controls and in TI vs TM patients. Easy ranking and identification of genes that are significantly down- or up-regulated was possible. The data provided genes that might be used as biomarkers in therapeutic management of Thalassemia patients.

D2.5 – Protein collection completed of at least 50 ErPC for WP2

Executive Summary
87 protein samples were collected from patients with β-thalassaemia (intermedia, TI and major, TM) and healthy controls, and these samples were inserted in the THALAMOSSprotein- Biobank. Following collection of samples, pilot iTRAQ experiments were performed to analyse and quantitate the samples from β-thalassaemia patients for globins levels and other known key erythropoietic factors. This analysis verified whether the protein samples in the THALAMOSS-protein-Biobank are suitable for proteomics profiling. The data obtained clearly indicate sharp differences between the thalassaemic and healthy samples showing the expected globin levels and the presence of known erythropoietic transcription factors. Thus, the erythroid differentiation protocols, protein

extraction method and the iTRAQ labelling/mass spectrometry will allow easy ranking and identification of proteins which are down- or up-regulated in ErPCs of thalassaemic versus healthy samples and TI versus TM samples. Thus, we conclude that the samples collected and the protocols used are suitable for proteomics analysis. Proteomics analysis will lead to identification of novel biomarkers for stratification of patients.

D2.6 – Mass spectrometry analysis completed of at least 50 protein samples

Executive Summary

52 protein samples from the Thalamoss Protein data-bank were collected from patients with β-thalassemia (intermedia, TI and major, TM) with respective healthy controls and were processed for iTRAQ labeling and mass spectrometry. The analysis is fully completed and the data obtained clearly indicated significant differences between the thalassemic and healthy samples, as well as between TI and TM samples. Thus, ranking was possible, as well as identification of proteins down- or up-regulated in ErPCs obtained from thalassemic patients versus healthy samples, and TI versus TM patients. Differentially expressed proteins highlighted changes in erythropoietic circuits between the different groups, and clarified mechanisms of erythopoiesis and the molecular differences between the different phenotypes of β-thalassaemia. The identified factors could be used for the identification of potential therapeutic targets or biomarkers for prognosis and therapeutic management of patients.

D2.9 – Functional analysis completed

Executive Summary

Based on the proteomics analysis, 17 proteins were selected for investigation of their effect on γ-globin expression in primary human erythroid cultures using shRNA technology. Among them successful repression of gene expression for twelve genes was shown by quantitative real-time PCR. In the absence of decitabine, reduced expression of the genes under investigation was accompanied by either a reduction or no change in γ-globin gene expression in primary human erythroid cultures. The only exception was gene 5, whose reduction resulted in an up-regulation of γ-globin gene expression at both the mRNA and protein levels in the absence of decitabine. Moreover treatment with decitabine in the presence of the shRNAs resulted in a significant increase in γ-globin gene expression in gene 9 and gene 11 knock-down cultures. As far as other HbF modifiers, ASH1L is a possible candidate, KLF1 was confirmed, and novel modifiers were proposed, such as arginine methyltransferases (PRMTs), their adapter molecules (BTGs), one of their targets (CHTOP), and members of the CCAAT-box binding transcription factor family C/EBP. Finally, a further possible HbF modifier was identified in LYAR (Ly-1 antibody reactive clone). With respect to this field of investigation, we have described a novel polymorphism (+25, G→A) of the Aγ-globin-gene leading to a decreased binding efficiency of nuclear factors to a LYAR-binding sites.

D2.11 – Globin monoclonal antibodies tested in clinical

Executive Summary

The major conclusion of D2.11 is that it is possible to isolate β-globin specific HCAb functional in clinical samples. After comparison with a commercially available anti-β-globin

antibody, we concluded that Harbor antibodies 2C10, 8A10 and 10H3 are good candidates for further development of β-globin assays.

D2.12 – Disease modifiers immunizations completed

Executive Summary
A number of heavy chain only antibodies have been isolated using peptide immunisations in the proprietary Harbour Antibodies mice. Immunisations to derive antibodies to human gpr 56, gpr 96 and gpr 114 (expressed on human stem cells and required for proliferation) was carried on.

D3.1 – 1st Report on HbF Inducers

Executive Summary
In D3.1 several novel HbF inducers were identified and several known HbF inducers further characterized. For instance we have investigated the effects on γ-globin genes of novel compounds designed starting from the structure of butyric acid. We have reported studies on the mechanism of action of mithramycin, sustaining the possible inhibitory effects on the molecular interactions between the Sp1 transcription factor and the promoters of the raptor and BCL11A genes. The studies described in D3.1 allow to propose microRNAs as possible players associated with high HbF production. As far as possible HbF inducers of therapeutic interest studies on rapamycin (Sirolimus) were included in D3.1. The results indicated that: (a) Sirolimus increases HbF in cultures with different basal HbF levels; (b) Sirolimus increases the overall Hb content/cell; (c) Sirolimus selectively induced γ-globin mRNA accumulation, with only a minor effect on β-globin and no effect on α-globin mRNAs; (d) there is a strong correlation between the increase by Sirolimus of HbF and the increase in γ-globin mRNA content. Other HbF inducers comparative studied were resveratrol (nine derivatives tested in K562 cells, one of which exhibiting an hemoglobin inducing activity comparable to the parent compound, being less cytotoxic than resveratrol), lenalidomide, angelicin, 5-aza-2'-deoxycytidine (decitabine), and mithramycin. In addition to HbF inducers, we attempted to inhibit globin gene expression. This might be important, since clinical complications in β-thalassemia are related also to accumulation of free globins which are not organized in a functional tetramer (such in the case of free α-globins in β-thalassemia). Inhibition of globin gene expression was obtained with antisense peptide nucleic acids (PNAs). Within D3.1 we also included the development of novel screening system for identification of HbF inducers, the human erythroid cell line HUDEP2 and K562 clones expressing different levels of BCL11A-XL.

D3.3 – 1st Report on HbF Inducers with Read-through Reagents

Executive Summary
We have constructed a yeast-based *in vivo* dual fluorescence read-through system that proved to be extremely simple and robust in the screening for the effect of drugs at each stop codon UGA, UAG and UAA in the same nucleotide context. In addition, we have provided evidence that the read-through response to low molecular weight drugs in yeast is essentially consistent with recent findings in human cells. Since the β039 thalassemia mutation is able to activate the nonsense-mediated decay (NMD) which recognizes and partially degrades nonsense transcripts, with subsequent rapid degradation of the β039

mRNA, the combined action of read-through molecules with the inhibition of the decay mechanism was proposed as a strategy to increase the amount of transcripts target+ and then the protein production. For this reason, we produced both clones with the NMD activated (NMD+) and clones with the decay inhibited (NMD-). Then we have produced human cellular clones carrying gene reporter sequences (green fluorescence protein, GFP) with stop codon mutations useful for a rapid and efficient molecular screening system of potential read-through activity. We have obtained a lentivirus vector with GFP gene carrying nonsense mutations by site specific mutagenesis and used it for transfection experiments in human cell line. We have also demonstrated this cellular model is suitable for the screening of molecules with read-through activity, in a 24-well plate format and analyzing GFP expression by FACS. Finally we have developed an experimental strategy based on siRNA to decrease the cellular content of the NMD-regulating proteins UPF1 (one of the most important factor involved in NMD process) in order to modulate NMD mechanism. The obtained clones were characterized in term of mRNA and protein and finally tested with read-through molecules, suggesting these further models for readthrough activity molecular screening.

D3.5 – 1st Report on GT Correction

Executive Summary

Several goals have been obtained under D3.5. Previously characterized gene therapy vectors were employed and novel lentiviral vector for gene therapy have been developed. The ability of various lentiviral constructs to improve adult (HbA) or fetal (HbF) hemoglobin synthesis was verified on 39 samples of the THALAMOSS biobank (see D1.7). The following lentiviral vectors have been used: AnkCT9W, AnkCT9WT87Q and novel unpublished constructs derived from small modifications of AnkCT9W. In addition, contructs able to increase HbF in adult erythroid cells. In these experiments healthy CD34+ cells isolated from peripheral blood were expanded, lentivirally infected with a construct carrying a ZF-Ldb1 transgene named GG1-SA, and differentiated along the erythroid pathway. GG1- SA expression led to a dramatic increase in the percentage of HbF compared to uninfected controls. Another field of investigation was the strategy based on combining gene therapy and treatment with fetal hemoglobin induction for the development of therapeutic protocols for β-thalassemia. To the aim of optimizing GT, duplex quantitation of LV vector copy number was developed and comparison of lentiviral vectors is ongoing. Finally, high efficiency mutation-specific gene-editing approach for the common IVS I-110 β- thalassaemia mutation was established.

D4.3 – Updated THALAMOSS Data Management Platform

Executive Summary

This Deliverable D4.3, the Updated THALAMOSS Data Management Platform, provides a fundamental update to the initial THALAMOSS Data Management Platform Deliverable D4.2, which has been generated as a prototype accompanied by extensive report describing its architecture and implementation. Current deliverable describes only major updates and operational activities that have been implemented since the initial release and tries to avoid duplication of the information, hence referring the reader also to the D4.2.
Based on additional requirements of the clinicians and results of biochemical analyses (WP1) as well as production of omics data (WP2), the data structure of the platformhas been substantially extended. The Updated THALAMOSS DataManagement Platform

includes finer-grained representation of genotyping, enhanced and restructured clinical parameters, support for raw and processed omics data with particular attention to genomics and proteomics, as well as new parameters representing results of various types of laboratory analyses. All the analytical results support also information about the specific analytical methods used for generating the data, in order to ensure reproducibility of the results, to support validation of measurements, and to support semantically correct harmonization of results.
Because of initial data collection period, which required extraction of data from various storage facilities without Internet access, has been considered mostly completed by the THALAMOSS partners, the updated version of the THALAMOSS Data Management Platform no longer supports the Microsoft Access user interfaces for offline data collection. This interface was very instrumental in the initial collection but turned out also very costly to support from the long-term perspective because of technology inconsistencies between Microsoft Access and common SQL databases, and because of ongoing substantial updates of the whole platform. Hence the Updated THALAMOSS Data Management Platform received extensive reimplementation of the web-based user interface, which follows all the updates of the data model as well as it features full data editing capabilities. It also has updated module for export of data for bioinformatics analyses.
From the operational perspective, the data curation and continuous data migration across the updates of the database constituted substantial part of the consumed resources. Another operational component was support of semi-automated import of the data from the contributors that already had large databases-the most outstanding example being data import from CING. As of submitting this deliverable, the THALAMOSS Data Management Platform exceeds 600 participants (patients) for which data is already entered.
Within the last year of the project, it is expected to add additional participants, as well as to extent the existing ones with more data, primarily results of ongoing omics analyses.

D4.4 – Bioinformatics Processing Software Suite

Executive Summary

Typical use of databases in production environments relies on carefully designed queries solving reporting tasks designed in advance. In a research environment, ad-hoc access to deeply structured relational databases is often required, where advance manual preparation of queries becomes problematic. We designed a system for automatic or semi-automatic query preparation to achieve data denormalization.
The system is implemented as a Python library, providing functions necessary to identify relations between variables and translating this information into a valid SQL query, executing this query and returning a flat file table suitable for downstream processing by statistical, machine-learning, visualization and other software. Such software often operates on flat tables rather than structured data. Consequently, library users only need to know (or select from a list) desired variables. After some calculations they receive an SQL command or a table of denormalized data reflecting the database content at the moment.
The denormalization procedure is parametrized and specific aspects of data manipulation can be controlled by the user. This includes data aggregation and formatting.
The Python library described here was tested on the first versions of THALAMOSS database and employed in a prototype web interface for data visualization. the web interface allows users to choose database variables and types of analysis/visualization to

be applied on the selected data. The data was also used in a number of machine-learning patient stratification analyses.

D4.6 – Suite of WEB applications provided on the ITHANET portal

Executive Summary

The ITHANET Portal offers a range of services including continuous updates on the upcoming conferences, interesting recent publications and talks. It also incorporates a database of organizations and networks concerned with haemoglobinopathies worldwide, and a wiki-based text interface to establish a growing, interactive education environment. The ITHANET Portal also houses one of the largest and most recent haemoglobinopathy mutation, frequency and genotype-phenotype databases, offering flexible search and display functionalities as a comprehensible and versatile reference for clinicians and researchers. The ITHANET Portal aims to become a comprehensive resource for all information relating to the haemoglobinopathies, and as an interactive community tool invites contributions to its content, including announcements, details of organizations and genotype-phenotype information. The ITHANET link to THALAMOSS is http://www.ithanet.eu/thalamoss.

During the first 36 months of the project the first two THALAMOSS web services has been developed and released on ITHANET web site: (a) the iron chelation prediction and (b) the HbF inducer prediction. ITHANET (http://www.ithanet.eu/thalamoss) will be constantly updated with new services and features until the completion of the project. The first web service is a classification model for the prediction of the Chelation outcome in thalassemia patients. The second tool aims to speed up the process of drug discovery by combining the principles of biology and chemoinformatics technology, allowing a more focussed approach to the task of screening for new HbF inducing agents. A Linux server has been developed with ability to export as a web service, with a GUI, every workflow developed in KNIME and extract data from every database or other web services with API.

The developed web applications for THALAMOSS can be accessed via the following links:

1. Prediction of chelation outcome
2. Prediction of HbF inducers (Chemoinformatics model)

D5.2 – Release II of the project website

Executive summary

The project's website constitutes the main point of collection and dissemination of the THALAMOSS project information, including public deliverables, summary of major scientific achievements, and advertisement of dissemination and training activities. Maintenance and update of the website take place on a weekly or monthly base. The private, password-protected section of the website is reserved for internal communication of the THALAMOSS Consortium, and it is used on a daily basis for internal communication. Furthermore, communication with EU and the Project Reviewers is possible through the password-protected section. At release II of the project website, the password-protected section of the website was updated and rearranged. The public pages are also updated with publications, executive summaries of the work, recent THALAMOSS meetings and training activities.

D5.10 – Establishment of patent portfolios around any technology/products

Executive Summary
Deliverable 5.10 is an overview of the possible THALAMOSS patent portfolios around the technologies and products that may arise from the project and have intellectual/commercial value. Both diagnostic and therapeutic areas of intervention have been considered, including drug repositioning issue and orphan drug designation. D5.10 is associated with D7.8 (IPR Management Database: Second Release).

D5.11 – Advertisement-promotion of the THALAMOSS Methodologies

Executive Summary
The THALAMOSS methodologies have been promoted by scientific publications on top Journals (including Blood, BMC Genomics, Molecular and Cellular Biology, Cell) as well as by more than 70 lectures and presentations. In addition THALAMOSS has been promoted at the Thalassemia International Federation Meetings held in Abu Dhabi (Dubai) and Athens (Greece).

D6.6 – Collection of the approvals from the Bioethic Committees on the use of biological materials from thalassaemic patients

Executive Summary
The WP6 deliverable D6.6 reports the overall activity of the THALAMOSS Consortium with respect to the submission to the Bioethic Committees of the THALAMOSS Project for approvals. The official approvals provided by UNIFE, KCL, LGHA, CING, CU and UNICA are included.

D6.8. Second THALAMOSS report of gender actions

ExecutiveSummary
In the first report of the Gender Action Plan for the THALAMOSS project (D6.7), we established the gender equality was maintained among partners and at all levels of the project. In this report, we will highlight women's contribution to data dissemination and support of the consortium for pregnant scientist. To extend further the action undertaken by the consortium it is proposed to organize a session for young female researchers to meet and discuss with successful female leaders at the next general assembly. The meeting will set the ground for the creation of mentorship for the women involved in the THALAMOSS project, it shall be beneficial for young researcher to benefit of guidance for next carrier move toward the end of the project.

D7.2. Second Periodic Project Report on Progress, Use of Resources and Financial Statement

ExecutiveSummary
In D7.2 is presented the Second Periodic Report on Progress, Use of Resources and Financial Statements.

D7.4 – Report on the Wider Societal Implications of the Project

Executive Summary
Deliverable D7.4 is related to describing the most important social impact of thalassemia and some relevant examples of the THALAMOSS activity in dissemination to the public and meeting social needs. THALAMOSS has already contacted several Thalassemia Patients' Associations, including the Cyprus Thalassemia Federation, the Cooley Anemia Foundation, the Thalassemia International Federation, the Veneta Association for the Fight Against Thalassemia, the Ferrara Association for the Fight Against Thalassemia. From all these organizations specific collaborations have been activated with the objective not only for disseminating THALAMOSS results, but also for collaborating in joint efforts to drive policy strategies on the field of rare diseases.

D7.8 – IPR Management Database: Second Release

Executive Summary
The technology and products developed in the THALAMOSS project during the second part of research activity are planned to be exploited in different ways by the various partners, mainly in the fields of discovery of novel therapeutic molecules for thalassemia and in-vitro diagnostics. This D7.8 deliverables presents the second analysis of the possible exploitable research within THALAMOSS.
