Associazione Italiana Sindrome di Shwachman-Diamond (AISS)
Via Pioveghetto 15, - 35136 Padova
Tel - FAX +39 049 8736130
E-mail: aiss@shwachman.it

Shwachman-Diamond Syndrome Italian Research Grant

Maximum Amount euro 10.000,00

Firm Deadline for Receipt of Applications: 31 December 2016

Eligibility: Persons applying for these grants if not in a faculty position need to provide a declaration by a supervisor with a position in the department (not a training position) and with authority to hold an independent research grant.

Terms of Support: Support may be provided for one (1) year in an amount not to exceed E 10,000. Indirect costs are permitted and are not to exceed 10% of the total costs.
The AISS will provide preference to those applications in which funds are used for supplies, equipment, technicians and other expenses and not for support of the salary of the PI or co-PIs.

Review: All applications will be reviewed by the AISS Scientific Committee (AISS-SC) or its designees.

Application: The application contains two sections.

Section 1, forms attached. The applicant and co-applicants must also include a current curriculum vitae.
Section 2: Research Plan, divided as indicated below. Parts A through D should not exceed 6 pages, using a font no smaller than 10 point.

Part A. Specific aims
Part B. Significance and background
Part C. Preliminary studies
Part D. Experimental design and methods
Part E. References (not included in the 6 page limit)
Part F. Relevance of the research to Shwachman-Diamond Syndrome
Part G. For junior faculty separate letter from supervisor or department head confirming commitment to project, and to provision of space and facilities
Part K. If human subjects and animals are involved, a statement by the PI or supervisor overseeing human or animal studies is compulsory. If considered as necessary by the AISS-SC, more information about ethical committee study approval may be asked.

Submission by email to the AISS: aiss@shwachman.it

1. Title of Proposal:

GENOMIC AND CLINICAL VARIATIONS IN SHWACHMAN-DIAMOND SYNDROME: IN SILICO AND MOLECULAR ANALYSIS OF DATA FROM EXOME SEQUENCING IN 16 PATIENTS.

2**. Applicant Information:**

Name:
ANTONELLA MINELLI

Title and Degree(s):
PhD

Work Address:
Università degli Studi di Pavia, Dip. di Medicina Molecolare, Unità di Biologia Generale e Genetica Medica, Via Forlanini 14, 27100 PAVIA

Telephone: + 39.0382.987727
Fax: + 39.0382.525030

Email: antonella.minelli@unipv.it

3. **Applicant Curriculum Vitae**: beginning on the next page, with 2 page limit. This will form application pages 3 and 4.

Curriculum Vitae

Personal data

Date and place of birth: Roma, July 1st, 1960. Nationality: Italian
Present position: Permanent researcher at Unit of General Biology and Medical Genetics
Institution: University of Pavia, Department of Molecular Medicine , Italy.

Education and training

- January 1984: degree in Biological Science at the University of Pavia, Italy.
- 1984: one-year training post-graduate at the Department of Hereditary and Human Pathology, University of Pavia, to be admitted to State Certification Exam for Biologist, obtained in 1985.
- 1985-1986: fellowship by "Associazione Italiana per lo Studio delle Malformazioni" (ASM) and post-graduate degree training at the Department of Hereditary and Human Pathology.
- May 1987: postgraduate Specialization in Human Cytogenetics, University of Pavia.
- 1988-1990: PhD training performed at the Department of Hereditary and Human Pathology in collaboration with the Department of Biology, University of Milano.
- June 1991: PhD in Human Pathology, University of Pavia.

Teaching experiences

From 1991 she has regularly performed teaching activity (lessons and tutorial activity), in courses of the Medical Faculty including the Schools enabled to provide degrees as Dietologist and Professional Nurse.
From 2001 to 2008 she taught in the Biology course in the Laurea degree in Physical Education.
From 2003 to 2014 she taught in the Biology and Medical Genetics courses in the Laurea degree in Nursing and Obstetrics.
From 2007 to 2010 she taught in the Human Genetics + Biology course in Master's degree of Biomedical Engineering.
From 2011 until now she teaches in the Medical Genetics and Prenatal diagnosis course in the Laurea degree in Biomedical Technician.

Scientific activity

The research activities are mainly focused on the molecular study of myeloproliferative syndromes, including the Shwachman syndrome (SDS) for its tendency in up to one third of cases, to develop myelodysplastic syndrome and acute myeloblastic leukemia; the parental origin of acquired chromosomal anomalies, as observed in haemathological disorders, was also studied.
The current research activity are concerned on the Shwachman syndrome and their topics are:

- mutation analysis of *SBDS* gene;
- study of clonal chromosome abnormalities by genotyping approach in bone marrow and in cells separated from peripheral blood (lymphocytes and granulocytes) to map the origin of chromosomal abnormality in the differentiation of hematopoietic tissue;
- study of cell biology focused on the morphology of neutrophils;
- Whole Exome Sequencing applied in a first group of SDS patients (Ref. 2015-AISS common project submitted by the scientific committee of the AISS).

As collaborator, she obtained funds from

1. Fondazione Cariplo for the project "Analisi genetiche e molecolari per la diagnosi e prevenzione di malattie rare: M. di Rendu-Osler-Weber (ROW) e S. di Shwachman (SS)" Ref.2002.2095/10.8485-08 ATTI 0002- PF RICERCA SCIENTIFICA 2002);
2. Fondazione Cariplo for the project 2012-0529.

As principal investigator, she obtained a Research Grant from Associazione Italiana Sindrome di Shwachman (AISS-2011).

From 2003 she provide molecular analysis of *SBDS* gene and her laboratory is present in the Orphanet database (http://www.orpha.net).

Published Articles inherent in SDS

1. Nacci L, Valli R, Pinto RM, Zecca M, Cipolli M, Morini J, Cesaro S, Boveri E, Rosti V, Corti P, Ambroni M, Pasquali F, Danesino C, Maserati E, Minelli A.Parental origin of the deletion del(20q) in Shwachman-Diamond patients and loss of the paternally derived allele of the imprinted L3MBTL1gene. Genes, Chromosomes & Cancer 2017;56:51–58.

2. Minelli A, Nacci L, Valli R, Pietrocola G, Ramenghi U, Locatelli F, Brescia L, Nicolis E, Cipolli M, Danesino C. Structural variation in *SBDS* gene,with loss of exon 3, in two Shwachman-Diamond patients. Blood Cells, Molecules and Diseases.2016;60:33–35.

3. Morini J, Babini G, Mariotti L, Baiocco G, Nacci L, Maccario C, Ro¨ßler U, Minelli A, Savio M, Gomolka M, Kulka U, Ottolenghi A, Danesino C. Radiosensitivity in lymphoblastoid cell lines derived from Shwachman–Diamond syndrome patients. Rad.Prot. Dosimetry. 2015; 166(1–4), 95–100

4. Nacci L, Danesino C, Sainati L, Longoni D, Poli F, Cipolli M, Perobelli S, Nicolis E, Cannioto Z, Morini J, Valli R, Pasquali F, Minelli A. Absence of acquired copy number neutral loss of heterozygosity (CN-LOH) of chromosome 7 in a series of 10 patients with Shwachman-Diamond syndrome. Br J Haematol 2014;165:573-575.

5. Minelli A, Nicolis E, Cannioto Z, Longoni D, Perobelli S, Pasquali F, Sainati L, Poli F, Cipolli M, Danesino C. Incidence of Shwachman-Diamond syndrome. Pediatr Blood Cancer. 2012;59:1334-1335.

6. Necchi V, Minelli A, Sommi P, Vitali A, Caruso R, Longoni D, Frau MR, Nasi C, De Gregorio F, Zecca M, Ricci V, Danesino C, Solcia E. Ubiquitin-proteasome-rich cytoplasmic structures in neutrophils of patients with Shwachman-Diamond syndrome. Haematologica 2012;97:1057-1063.

7. Maserati E, Pressato B, Valli R, Minelli A, Sainati L, Patitucci F, Marletta C, Mastronuzzi A, Poli F, Lo Curto F, Locatelli F, Danesino C, Pasquali F. The route to development of myelodysplastic syndrome/acute myeloid leukaemia in Shwachman-Diamond syndrome: the role of ageing, karyotype instability, and acquired chromosome anomalies. Br J Haematol. 2009;145:190-197.

8. Minelli A, Maserati E, Nicolis E, Zecca M, Sainati L, Longoni D, Lo Curto F, Menna G, Poli F, De Paoli E, Cipolli M, Locatelli F, Pasquali F, Danesino C. The isochromosome i(7)(q10) carrying c.258+2t>c mutation of the *SBDS* gene does not promote development of myeloid malignancies in patients with Shwachman syndrome. Leukemia. 2009; 23:708-711.

4. **Applicant's Commitment as Investigator of the Project**:

I agree as the applicant to accept responsibility for the scientific management of this project as outlined in this application. I further agree to submit a report at the end of the granting period.

5. **Applicant's Affirmation**:

I certify that the investigations involving human subjects to be carried out in the application will have approval of the applicant's Institutional Ethical Committee

Approvals from the Institutional Ethical Committee must be included with the application.

6. **Research Results**:

Results of research may be made available to the public through appropriate scientific channels. All publications will bear the statement:

Antonella Minelli
Signature of Applicant

23/12/2016
Date

7. Applicant's Institution Certification and Commitment:

I certify that the statements herein and the Applicant's Affirmation are true, complete and accurate to the best of my knowledge and I agree to accept responsibility for the fiscal management of this project as outlined in this application. I further agree to commit this institution to comply with the Associazione Italiana Sindrome di Shwachman-Diamond (AISS) terms and conditions if a grant is awarded as a result of this application.

Name of Institution Official:
Prof. Ermanno Gherardi

Title:
Full Professor of GENERAL PATHOLOGY

Address:
Dipartimento di Medicina Molecolare,

Phone: +390382986845
Fax: _____

E-mail Address: egherard@unipv.it

Signature of Institution Official

23/12/2016
Date

ABSTRACT OF RESEARCH PLAN

Within the space provided, summarize the long-term objectives, scientific aims and methodology of the proposal.

TITLE: GENOMIC AND CLINICAL VARIATIONS IN SHWACHMAN-DIAMOND SYNDROME

Objectives and aims. The project is intended to (partially) explain why in patients affected with Shwachman-Diamond Syndrome (SDS) the same common mutations, in the only known disease causing gene, *SBDS*, are associated to a wide clinical heterogeneity. In fact, three mutations: c.183_184TA>CT; c.258+2T>C; and c.[183_184TA>CT; 258+2T>C] account for more than 75% of mutated alleles.

Background/rationale. SDS (OMIM#260400) is an autosomal recessive bone marrow failure syndrome. It is characterized by exocrine pancreatic dysfunction, haematologic abnormalities, bone marrow (BM) failure, neurodevelopmental delay and skeletal abnormalities. All the listed clinical problems are observed only in part in the patients, even if they share exactly the same mutations; clinical heterogeneity is observed also among sibs. Thus, it is likely that genetic variations in genes other than *SBDS* will act as additional modifier factors to explain the differences in observed clinical phenotype in different patients.

Research design. The project will perform a bioinformatic analysis on the data obtained by whole exome sequencing applied in 16 patients; the exome sequencing was previously funded by AISS.

This analysis will have as target a set of genes known to be relevant for well-defined clinical problems as pancreatic function, growth and skeletal development, bone marrow morphological and cytogenetic abnormalities, cognitive impairment.

The bioinformatics analysis will be followed by laboratory confirmation of the variants of interest in parents to discriminate "de novo" from inherited variants and in a larger number of cases.

Anticipated output. For each clinical problem, we expect to identify a small number of DNA variants, common to patients with a severe clinical presentation, but absent in patients with a mild or absent expression of the selected clinical problem. These variants, when added to the background of biallelic *SBDS* mutation, will modify the clinical presentation of the disease.

The relevance of the expected results, is that we believe to be able to start to identify for SDS, a "genomic picture" far more complex and informative than the "simple" presence of mutations in the disease causing gene.

BUDGET

List below a budget by categories for the support. The review committee will carefully consider the appropriateness of your budget. It must be well defined, justified, and realistic to complete the work proposed. The first column defines the total expenses that are expected to be necessary to realistically complete the project. The second column indicates the expenses requested from the AISS. Applicants **will not** be penalized in funding considerations for requiring additional funds beyond what is requested from the Foundation(AISS); however, the true costs of the project must be acknowledged.[This and the section on page 1 re: Other Funding need to be consistent]

EURO Amount Requested for:

	TOTAL COSTS REQUIRED TO COMPLETE PROJECT:	**COSTS REQUESTED FROM AISS**:(not to exceed E 10,000)
Personnel (including fringe benefits): PI: Name: Dr. Antonella Minelli Co-I Name: Prof. Cesare Danesino Additional personnel (identify role): A fellowship for 1 year Name: Dr. Lucia Nacci	25,000 euros/year	No request
Equipment: All needed equipment is available	---	---
Supplies: Laboratory supplies (consumables, expenses for sequencing)	15.000/year	7.500/year
Other Expenses: General expenses for carrying on the project, bioinformatics expenses	10.000/year	2.500/year
Indirect Costs (not to exceed 10% of total)	1.000/year	
TOTAL COSTS:	51.000/year	10.000/year

Justification: Define and justify expenses in each category. Explain the role of each of the individuals named in the Personnel section. The justification must include an explanation of what each category contributes to the project. Also explain any marked differences between the first- and second-year expenses in a particular category. The AISS will provide preference to those applications in which funds are used for supplies, equipment, technicians and other expenses and not for support of the salary of the PI or co-PIs. The AISS-SC may ask for further expense details.

The total cost of the project , over a two year working time is 102.000Euros, 51.000 euro /year.
A salary for a fellowship for a post-Doc is justified by the large amount of in silico work, which certainly will need full time availability of one person.
The following work of validation of the variants and of extension of the analysis of selected variants to a larger set of cases will require 40% of available working time of the PI, 10 % of available working time of the Co-I.

Other expenses include: laboratory consumables (DNA extraction Kit; DNA purification columns; custom sequencing costs; pipet tips, tubes); reagents for DNA amplification.
General expenses will include all expenses related to the use of dedicated bio-informatic programs, general expenses for running the laboratory, including secretarial expenses and general expenses of the Department, including overheads.

Other Support for this Project:

Applicants are allowed to receive funding from other sources for parts of the project not funded by the AISS. Please, list all other funding sources.

Additional funding needed to complete the project will be received from University of Pavia (salaries to PI and Co-I), and other requests to research dedicate charities are pending.

Research Plane

Part A. Specific aims

The project is aimed to correlate genetic variation and clinical heterogeneity in Shwachman-Diamond Syndrome (SDS) by analysing the results of Whole-Exome Sequencing (WES) already performed in 16 cases with a bioinformatic analysis on the data, and by Sanger confirmation of relevant DNA variation.

Objectives of the project are: i) to identify variants in genes other than the disease causing gene *SBDS* able to modify the final clinical phenotype and to act as risk factor for the more relevant clinical problems observed in SDS patients; ii) to use these variants for better prognostic evaluation of the course of the disease in youngest patient; iii) to use our results as a tool to apply personalized medicine in SDS.

Thanks to the extensive collaboration with almost all Italian Haematology and Gastroenterology centres, we has been able to confirm the clinical diagnosis in over 100 Italian patients with molecular analyses. Therefore, a large amount of data about clinical pictures and follow-up of mutation confirmed patients is already available.

Part B. Significance and background

Shwachman Diamond Syndrome (OMIM#260400) is an autosomal recessive bone marrow failure syndrome related to mutations in the *SBDS* gene (Boocoock *et al.,* 2003).

SDS incidence was estimated as 1 in 76,563, and Minelli et al. in 2012 reported a much lower incidence in Italy (1: 168,000) (Minelli *et al.,* 2012); up to December 2016, the Italian SDS Registry for the disease includes 120 cases. SDS shows a wide range of abnormalities and symptoms including exocrine pancreatic dysfunction, haematologic abnormalities (Mack *et al*., 1996) bone marrow (BM) failure (Donadieu 2012), neurodevelopmental delay and skeletal abnormalities (Mäkitie *et al*., 2004). Additionally, patients may show abnormalities in many other organs and systems, most notably in the bones, liver, heart and immunologic system. Overall, the clinical phenotype shows a wide variability among patients. Dror (2011) included the demonstration of *SBDS* mutations among the diagnostic criteria.

One of the hallmarks of SDS is exocrine pancreatic dysfunction of varying severities caused by absence of acinar cells; it is usually diagnosed within the first six months of life and (in 90% of cases) during the first year (Mack *et al*., 1996). Patients classically present in early infancy with malabsorption, steatorrhea, failure to thrive, and low levels of fat-soluble vitamins A, D, E, and K (Burroughs *et al*., 2009). The most common hematologic abnormality affecting over 80% of patients with SDS is neutropenia, typically defined as a neutrophil count less than 1,500 × 109/L, but the hematopoietic defect can also affect red cells and platelets (Huang and Shimamura, 2011). SDS patients are at increased risk of developing bone marrow aplasia, myelodysplastic syndrome and/or acute myeloid leukaemia (MDS/AML); the risk is higher than in general population, and 15%-25% of SDS patients will develop serious haematological problems (Dror *et al*., 2002; Burroughs *et al*., 2009; Donadieu *et al.,* 2012).

Skeletal abnormalities are also common, and include skeletal dysplasia (Ginzberg *et al.,* 1999; Makitie *et al.,* 2004), low turnover osteoporosis, vitamin D deficiency independent (Toiviainen-Salo *et al.,* 2007). By the 1st birthday, over half of patients drops below the 3rd percentile for both height and weight.

The neuropsychological phenotype of SDS patients is still poorly characterized, but impaired psychosocial functioning and various degrees (from mild to severe) of cognitive impairment are present in the majority of patients (Perobelli *et al*., 2012).

Perobelli *et al.*(2012), using a multi-modal approach including general and specific neurocognitive tests and structural MRI, hypothesized that cognitive impairment in SDS patients is associated with diffuse brain anomalies in the grey matter and white matter connectivity.

Boocock reported that *SBDS*, was mutated in SDS patients (Boocock *et al.,* 2003). This gene contains five exons encompassing 7.9 kb and maps to the 7q11.21 centromeric region of chromosome 7 (Goobie *et al*., 2001).

Up to now, *SBDS* is the only gene associated with SDS; molecular testing is routinely available, and three

mutations (resulting from gene conversion between *SBDS* and SBDSP) account for more than 75% of mutated alleles: i) c.183_184delinsCT; ii) c.258+2T>C; iii) c.[183_184delinsCT; 258+2T>C]; in the latter, both DNA changes occur on the same allele. If in a child i) and ii) are found, parents need to be tested to determine whether the mutation(s) observed in their child are monoallelic (as in iii) or biallelic. Targeted mutation analysis is usually performed as this strategy detects at least one mutation in over 90% of patients.
About 85% of patients are compound heterozygous for the two common mutations; direct sequencing of the five exons of the gene detects less common mutations, including missense, nonsense, insertions, and deletions. The two recurrent mutations in exon 2 insert new cutting sites for two restriction enzymes. The c.[183_184TA>CT] nonsense mutation results in a stop codon (p.K62X). The c.[258+2T>C] mutation causes a splicing error with a deletion of 8 bp and result in a frameshift (p.84Cfs3). The c.[258+2T>C] mutation still allows the production of some amount of normal protein (Austin *et al.,* 2005), both in heterozygotes and homozygous (Nicolis *et al.,* 2007).
SDS patients have a propensity to acquire characteristic clonal chromosomal changes in the BM, mainly involving chromosomes 7 and 20. The most frequent anomalies are an isochromosome for the long arms of chromosome 7, i(7)(q10) (Pressato *et al.,* 2010), and an interstitial deletion of the long arms of chromosome 20, usually del(20)(q11.21q13.32) (Maserati *et al*., 2006; Valli *et al*., 2013).
Minelli using microsatellite analysis, demonstrated that the i(7)(q10) was always derived from the parental chromosome containing the splicing mutant allele c.[258+2T<C] (Minelli *et al.,* 2009). As this mutation still allows the production of some amount of normal protein (Nicolis *et al.,* 2007), these results suggest a selective advantage of cells having two copies of an allele allowing partial expression of *SBDS*, and persistence and expansion of the clones containing the i(7)(q10) (Minelli *et al.,* 2009).
The second common cytogenetic abnormality seen in SDS patients is an interstitial deletion of the long arms of chromosome 20, del(20)(q). A possible explanation of the low risk of MDS/AML in patients carrying del(20)(q) (as the only clonal abnormality), was provided by Pressato *et al.* (2012) and Valli *et al.* (2013) who demonstrated the loss of *eIF6* gene, mapping in the deleted region of chromosome 20. This gene has a pivotal role in ribosome biogenesis together with SBDS and its partial loss concomitant to the absence of SBDS protein may play a specific role to lower the risk of transformation into MDS/AML in SDS patients.
Evidence of a link between SBDS and ribosomal function or RNA metabolism comes from the presence in bone marrow cells from SDS patients of abnormal expression of genes involved in ribosome biogenesis, rRNA and mRNA processing, together with a decreased expression of several ribosomal proteins involved in cell growth and survival (Rujkijyanont *et al.,* 2009). In conclusion, SBDS is a multifunctional protein involved in ribosome biogenesis, and Shwachman-Diamond syndrome has been recently defined as a ribosomopathy caused by uncoupling GTP hydrolysis from eIF6 release during late citoplasmic maturation of pre-60S subunits.

Rationale

Based on data reported in background, it is clear that: i) SDS is a multisystem diseases caused by a very few number of mutations in the *SBDS* gene; ii) that all the more relevant clinical problems (bone marrow failure and risk for MDS or AML, skeletal dysplasia, cognitive impairment) are present only in a variable percentage of patients; iii) that there is no correlation between the presence of any mutation and the risk for the patients do develop the afore mentioned clinical problems.
We can define the presence/absence of a defined clinical sign as extreme phenotypes. In order to try to identify the genetic factors acting as modifiers of the clinical phenotype, the best approach is to analyze all DNA variants present in each patient and parents (trios).
As we have now the exome data from 16 patients, we can now compare cases showing the same extreme phenotype (severe BM failure, same example as before), and we expect them to share a number of variants, which are conversely not present in cases in whom, for instance, BM is normal or minimally affected.
The “extreme phenotype” design was proven useful for instance for Cystic Fibrosis, with the demonstration that variants in *DCTN4* may influence P. Aeuriginosa infection in cystic fibrosis, and recently a partially

modified approach, (i.e., extreme phenotype vs. control population design) allowed the identification of additional variants in CAV2 and TMC6 (Emond *et al.*, 2015).
Analysis of the shared variants in the whole group of 16 cases will enable us to recognize groups of cases with extreme phenotype for all the above mentioned relevant clinical problems, as already done in the first group of 8 patients (see preliminary data).
As DNA is available for the vast majority of case included in the Italian Registry (n=120), as well as essential clinical data, we expect to be able to confirm the relevance of variants identified in small group of cases with each extreme phenotype in a larger group of cases.

Part C. Preliminary studies

All cases studied have been entered by M. Cipolli in the Italian Registry, (http://www.registroitalianosds.org), and the Co-I is a component of its Scientific Committee. Preliminary results include a complete clinical workout of each of the patients (including hematology, gastroenterology, paediatrics, cognitive evaluation), molecular analyses, bone marrow cytogenetics, genetic counselling.
In the registry are included 6 families with more than 1 affected child, 3 of them are under regular follow up. Clinical data are available demonstrating relevant differences among sibs, and one of these families is included among the 166 patients studied.
For all patients in whom a clonal BM abnormality [either i(7)(q10) or del(20)(q)] was found the parental origin of the anomaly was assessed by studying the relevant STRs.
The AISS (Associazione Italiana Sindrome di Shwachman) in 2015 funded the WES analysis on 16 cases, so all the raw data about Whole-Exome Sequencing in these cases are already available.
The 16 SDS patients studied listed in Table 1, which indicates that 10 of them show the commonest genotype, while in 6 cases uncommon mutations were observed.
We obtained from Personal Genomics the full list of variants observed, filtered only to exclude synonymous variants, and in Table 2 are entered the number of variants observed for each case.
In silico analysis was started by selecting in the KEGG PATHWAY Database the following pathways:

1) Cell cycle (including 124 genes), because of previous work by our group (Morini et al, 2015; experiments performed by L. Nacci at University of Cambridge);
2) Pancreatic secretion (including 96 genes), because pancreatic insufficiency is one of the major clinical sign of SDS.
3) Ribosome (including 137 genes) and
4) Ribosome biogenesis in eukaryotes (including 87 genes), as SDS is a ribosomopathy.

From the general database, including all the DNA variants identified in the 8 patients, we selected first the variants present in the genes related to each pathway and present in the 8 patients; the percentage of genes involved for each pathway in the 8 SDS patients is entered in Table 3. The differences in percentage of genes carrying variants for each pathway is statistically significant (χ^2-test: p=0.00006).
Next, for each pathway and for each of first 8 patients, we selected the variants predicted to be damaging in at least 3 out of 5 Functional Prediction Databases (SIFT, PolyPhen2, Mutation Taster, Mutation Assessor and FATHMM) (Table 4).
For each pathway, we observed that small groups of 2-3 cases were similar for the presence of a smaller or larger number of variants. In Table 5, we entered the details about cell cycle pathway, with the details of genes carrying the variants selected, as above indicated.
Similar data are already available for all the pathway previously quoted (data not included).

We also performed hierarchical clustering analysis to evaluate the similarities among single nucleotide variation (SNV) founded in SDS patients for each pathway of interest. In Figure 1a we report an example of this type of analysis, and similarly the same data are already available for all pathways for the first group of cases. The cluster dendrogram (Fig.1b) reflects how similar the patients are: UPN 20 is different from all other

patients, while UPN 6 and UPN 58 are quite similar, forming a subgroup far from other patients.
Overall, preliminary results indicate that the data obtained by WES can be used to identify small groups of patients (with the same genotype as far as *SBDS* mutations are concerned) in whom similarities for the presence of a number of DNA variants are likely to (partially) explain the clinical variability.

Part D. Experimental design and methods

Bioinformatic analysis of WES results

As summarized in Preliminary Results, we aim to perform this type of analysis in the whole group of 16 cases. To assess the significance of the variants identified, if pathogenic or not, the following bioinformatics tools will be used:
- SIFT (https://www.sift.jcvi.org/), predicts whether an amino acid substitution affects protein function, and is based on the degree of conservation of amino acid residues in sequence alignments derived from closely related sequences;
- PolyPhen2 (https://www.genetics.bwh.harvard.edu/pph2/), is a tool which predicts possible impact of amino acid substitution on the structure and function of a human proteins using physical and comparative considerations;
- Mutation Taster (https://www.mutationtaster.org/).
The results provided by the tools listed are based on different strategies, so it is possible to observe conflicting result if a single variant is tested by all of them. We will first take into account variants considered as pathogenic by all three tools. We will use the OMIM database to link genes containing possible damaging variants to human diseases, and the ExAC Browser Beta (http://www.exac.broadinstitute.org/) to assess the frequency of variants in European population.

Pathway analysis

We will select some pathways relevant to SDS phenotype using the following tools:
- Kyoto Encyclopedia of Genes and Genomes (KEGG) PATHWAY Database (http://www.genome.jp/kegg/pathway.html);
- PANTHER classification system (http://www.pantherdb.org/pathway/), pathway commons (http://www.pathwaycommons.org/pc/)
Hierarchical clustering analysis of samples will be performed to evaluate the similarities among SDS patients' SNVs for the different KEGG pathways of interest. Euclidean distance matrices for each pathway were performed in R (http://www.r-project.org/) with the "stats" package *dist()* function using a binary scoring matrix to distinguish the presence, or not, of a SNV in each patients'exomes. Heatmap visualization of such distance matrices is obtained with the R package "gplots" and show the unique "fingerprints" of SNVs for each patient through a dual color representation: yellow if the variant was found while red if not. The dendrogram on top of the heatmap reflects the separation in observed SNVs among the different SDS patients.

WES result analysis

For each of the clinical problems selected, as a) growth retardation and skeletal dysplasia; b) bone marrow abnormalities and/or evidence of MDS; c) pancreatic function; d) cognitive impairment; e) presence of selected clonal bone marrow cytogenetic abnormalities as del(20) we will select, using the above listed pathway databases, a set of related genes.
For instance, for pancreatic function in the KEGG database we find the "fat digestion and absorption" pathway and the "pancreatic secretion" pathway, including respectively 41 and 96 genes. We will search the WES results of each case for variants present in the two group of genes, and referred as pathogenic in at least three out of the three prediction tools listed above, and for variants whose frequency in European population is below 3%. We will then split the cases according to "extreme phenotypes", UPNs with severe impairment of pancreatic function without spontaneous improvement vs UPNs with absent/mild pancreatic function impairment with spontaneous improvement.
Next, we will compare the two groups, to identify variants common to severe phenotypes, and absent in mild phenotypes. Hierarchical clustering analysis will identify subgroups of cases based on their similarities for the

presence/absence of variants (see preliminary results).
As we will analyse patients and parents, we will know which variants are *de novo*. For de novo variants, we will use the same prediction tools to select those predicted to be pathogenic. If some of them are found in a gene related to human diseases, we will ask the doctor in charge of the UPN in whom it was found to revise the clinical picture to verify if the phenotype is a complex one eventually including symptoms uncommon to SDS patients, which might be related to a second genetic disease. The "extreme phenotypes" for other clinical problems will be defined as follows:
- skeletal abnormalities and growth: height below 3rd centile or above 10th centile; presence or absence of severe skeletal dysplasia requiring orthopedic surgery;
- bone marrow abnormalities: severe hypocellularity and extensive fatty infiltration of the marrow compartments for age, indicating marrow failure and disordered hematopoiesis, MDS vs absent or mild bone marrow abnormalities;
- cognitive impairment: presence of cognitive impairment assessed by the questionnaire proposed by the Italian Institute of Statistics for their 2004 Enquiry on Disabilities (http://www.disabilitaincifre.it/documenti/salute.pdf), vs absence of cognitive impairement.
- Clonal bone marrow cytogenetic abnormality: presence of del(20) in the last BM control vs absence of the clonal abnormality.

Independently from the bioinformatics analyses listed above, we will perform a hierarchical clustering analysis on the whole group of cases, and cases clustering together, will be compared by the clinicians to record the clinical similarities.

Part E. References (not included in the 6 page limit)

Austin KM, Leary R.J, Shimamura A. The Shwachman–Diamond SBDS protein localizes to the nucleolus. Blood. 2005; 106: 1253–1258.

Boocock GR, Morrison JA, Popovic M, Richards N, Ellis L, Durie PR & Rommens JM. Mutations in SBDS are associated with Shwachman-Diamond syndrome. Nature Genetics. 2003; 33:97–101.

Burroughs L, Woolfrey A, Shimamura A. Shwachman-Diamond Syndrome: A Review of the Clinical Presentation, Molecular Pathogenesis, Diagnosis, and Treatment. Hematology/oncology clinics of North America. 2009; 23:233–248.

Donadieu J, Fenneteau O, Beaupain B, et al. Classification of and risk factors for haematologic complications in a French national cohort of 102 patients with Shwachman-Diamond syndrome. Haematologica. 2012; 97: 1312-1319.

Dror Y, Durie P, Ginzberg H, Herman R, Banerjee A, Champagne M, Shannon K, Malkin D, Freedman MH. Clonal evolution in marrows of patients with Shwachman-Diamond syndrome: a prospective 5-year follow-up study. Exp Hematol. 2002; 30:659-669.

Dror Y, Donadieu J, Koglmeier J, Dodge J, Toiviainen-Salo S, Makitie O, Kerr E, Zeidler C, Shimamura A, et al. Draft consensus guidelines for diagnosis and treatment of Shwachman-Diamond syndrome. Ann N Y Acad Sci. 2011; 1242:40-55.

Emond MJ, Louie T, Emerson J, Chong JX, Mathias RA, Knowles MR, Rieder MJ, Tabor HK, Nickerson DA, Barnes KC, Go L, Gibson RL, Bamshad MJ, Project NGES. Exome Sequencing of Phenotypic Extremes Identifies CAV2 and TMC6 as Interacting Modifiers of Chronic Pseudomonas aeruginosa

Infection in Cystic Fibrosis. PLoS Genet 2015;11:e1005273.

Ginzberg H, Shin J, Ellis L, Morrison J, Ip W, Dror Y, Freedman M, Heitlinger LA, Belt MA, Corey M, Rommens JM, Durie PR. Shwachman syndrome: phenotypic manifestations of sibling sets and isolated cases in a large patient cohort are similar. J. Pediatr. 1999; 135: 81–88.

Goobie S, Popovic M, Morrison J, Ellis L, Ginzberg H, Boocock GR, Ehtesham N, Bétard C, Brewer CG, Roslin NM, Hudson TJ, Morgan K, Fujiwara TM, Durie PR, Rommens JM. Shwachman-Diamond syndrome with exocrine pancreatic dysfunction and bone marrow failure maps to the centromeric region of chromosome 7. Am. J. Hum. Genet. 2001; 68: 1048–1054.

Huang J, and Shimamura A. Clinical spectrum and molecular pathophysiology of Shwachamn-Diamond syndrome. Carr Opin Haematol. 2011; 18: 30-35.

Mack DR, Forstner GG, Wilschanski M, Freedman MH, Durie PR. Shwachman syndrome: exocrine pancreatic dysfunction and variable phenotypic expression. Gastroenterology, 1996; 111:1593-1602.

Mäkitie O, Ellis L, Durie PR, Morrison JA, Sochett EB, Rommens JM, Cole WG. Skeletal phenotype in patients with Shwachman-Diamond syndrome and mutations in SBDS. Clin Genet. 2004; 65:101-112.

Minelli A, Nicolis E, Cannioto Z, Longoni D, Perobelli S, Pasquali F, Sainati L, Poli F, Cipolli M, Danesino C. Incidence of Shwachman-Diamond syndrome. Pediatr Blood Cancer. 2012; 59:1334-1335.

Nicolis E, Durie PR, Ip WF, Rommens J, Shimamura A, Sainati L, Di Meglio A, Danesino C, Zecca M, Assael BM, Rizzotti P, Cipolli M. SBDS mRNA Expression in Shwachman–Diamond Syndrome Patients Carrying the 258+2T>C Mutation. Proceedings of the 4th International Congress on Shwachman–Diamond Syndrome; Boston, MA. 10–12 June 2007.

Morini J, Babini G, Mariotti L, Baiocco G, Nacci L, Maccario C, Ro¨ßler U, Minelli A, Savio M, Gomolka M, Kulka U, Ottolenghi A, Danesino C. Radiosensitivity in lymphoblastoid cell lines derived from Shwachman–Diamond syndrome patients. Rad.Prot. Dosimetry. 2015; 166(1–4), 95–100.

Perobelli S, Nicolis E, Assael BM, Cipolli M. Further characterization of Shwachman–Diamond syndrome: psychological functioning and quality of life in adult and young patients. Am. J. Med. Genet. 2012; 158:567–573

Pressato B, Marletta C, Montalbano G, Valli R, Maserati E. Improving the definition of the structure of the isochromosome i(7)(q10) in Shwachman-Diamond Syndrome. Br J Haematol. 2010; 150:632-633.

Pressato B, Valli R, Marletta C, Mare L, Montalbano G, Lo Curto F, Pasquali F, Maserati E. Deletion of chromosome 20 in bone marrow of patients with Shwachman-Diamond syndrome, loss of the EIF6 gene and benign prognosis. Br J Haematol. 2012; 157:503-505.

Rujkijyanont P, Adams SL, Beyene J, Dror Y. 2009. Bone marrow cells from patients with Shwachman-Diamond syndrome abnormally express genes involved in ribosome biogenesis and RNA processing. Br J Haematol. 2012; 145:806-815.

Toiviainen-Salo S, Mäyränpää MK, Durie PR, Richards N, Grynpas M, Ellis L, Ikegawa S, et al. Shwachman-Diamond syndrome is associated with low turnover osteoporosis. Bone. 2007; 41:965–972.

Valli R, Pressato B, Marletta C, Mare L, Montalbano G, Curto FL, Pasquali F, Maserati E. Different loss of material in recurrent chromosome 20 interstitial deletions in Shwachman-Diamond syndrome and in myeloid neoplasms. Mol Cytogenet. 2013; 6:56.

Part F. Relevance of the research to Shwachman-Diamond Syndrome

Shwachmann–Diamond is an autosomal recessive disorder and most patients have the same genotype, as they carry two common mutation in the SBDS gene. Still the main clinical signs (neutropenia, pancreatic insufficiency, skeletal dysplasia, cognitive impairment, bone marrow abnormalities, clonal cytogenetic abnormalities, risk to develop AML) can be observed with a probability ranging from 98% (neutropenia) to 30% (risk for AML).

Thus a project aimed to identify additional DNA variants acting as modifying factors, and behaving as risk factors for the development of specific symptoms, is of relevance for patients and their families as it will offer improved tools to define the prognosis of the disease in each single case, and may be of relevance for family planning and prenatal diagnosis.

Identification of genetic variants additional to the mutations in the disease-causing gene, also might, in the future, have some relevance in better selection of therapies for specific clinical symptom.

TABLES AND FIGURES

Table 1: patients studied with their *SBDS* mutations

N. of case	ID case	*SBDS* mutations
1	UPN 20	258+2T>C/183_184TA>CT
2	UPN 6	258+2T>C/101A>T
3	UPN 68	258+2T>C/183_184TA>CT
4	UPN 57	258+2T>C/187G>T
5	UPN 58	258+2T>C/183_184TA>CT
6	UPN 15	258+2T>C/183_184TA>CT
7	UPN 65	258+2T>C/258+2T>C
8	UPN 64	258+2T>C/624+1G>C
9	UPN 42	258+2T>C/c.258+533_459+403del
10	UPN 43	258+2T>C/c.258+533_459+403del
12	UPN 2	258+2T>C/356G>A
13	UPN 62	258+2T>C/183_184TA>CT
14	UPN 85	258+2T>C/183_184TA>CT
15	UPN 24	258+2T>C/183_184TA>CT
16	UPN 51	258+2T>C/258+2T>C
17	UPN 45	258+2T>C/183_184TA>CT

Table 2: first group of 8 patients analyzed with the number of their variants

N. of case	ID case	n. of variants
1	UPN 20	9821
2	UPN 6	10005
3	UPN 68	9975
4	UPN 57	10091
5	UPN 58	10111
6	UPN 15	10168
7	UPN 65	10189
8	UPN 64	10038

Table 3: pathways analysed and % of genes involved in the 8 SDS patients.

Pathway	Genes in the pathway	Genes containing variation in the 8 SDS patients	%
Cell cycle	124	42	33.8
Pancreatic secretion	96	46	47.9
Ribosome	137	29	21.16
Ribosome biogenesis in Eukaryotes	87	39	44.82

Table 4: variants predicted to be damaging in at least 3 out 5 Functional Prediction Databases

	UPN 20	UPN 6	UPN 68	UPN 57	UPN 58	UPN 15	UPN 65	UPN 64
N. of variants in genes involved in cell cycle pathway	2	0	1	0	0	0	1	3
N. of variants in genes involved in pancreatic secretion pathway	2	2	3	6	3	5	7	2
N. of variants in genes involved in ribosome pathway	1	2	1	3	0	1	1	0
N. of variants in genes involved in cell cycle pathway	2	3	4	2	4	1	3	3

Table 5: details of variants selected as pathogenic, their frequency in European population and distribution in the 8 SDS patients in cell cycle pathway

chromosome: position	gene	identifier	UPN 20	UPN 6	UPN 68	UPN 57	UPN 58	UPN 15	UPN 65	UPN 64	Eur Freq.
11:108143456	ATM	rs1800057	x							x	0,026
12:69233376	MDM2	rs201788800								x	0,000
16:53488627	RBL2								x		0,000
16:53499374	RBL2	rs76818213								x	0,019
22:29091782	CHEK2	rs373073383			x						0,000
22:45749966	SMC1B	rs61735519	x								0,060

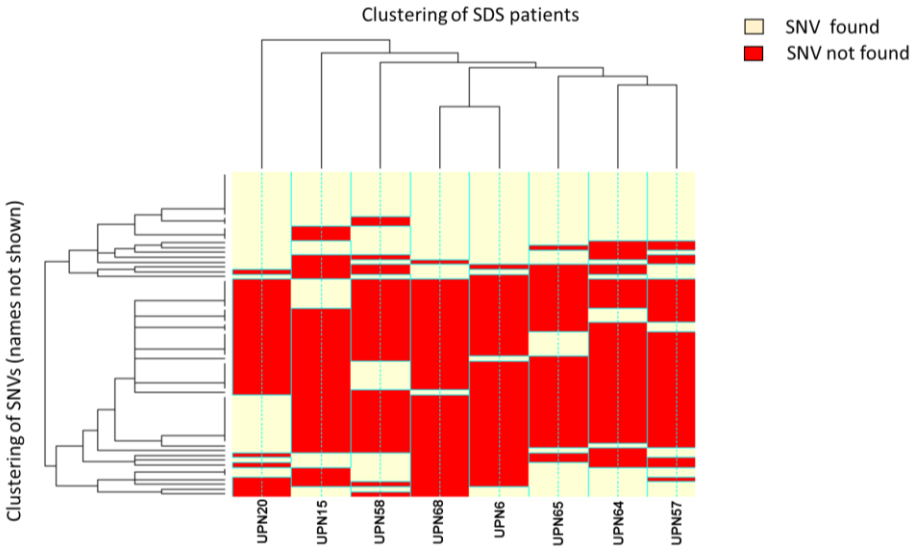

Fig 1a: Hierarchical clustering of SDS patients using a binary scoring matrix to distinguish the presence, or not, of SNVs in each patients' exomes in the cell cycle pathway. Each row and column refer to an individual SNV and an individual SDS patient, respectively. The yellow and red color reflect absence or presence of SNV, respectively. At the top of hierarchical clustering is reported the separation of SDS patients in different clusters based on the similarity or not of their unique SNVs "fingerprints".

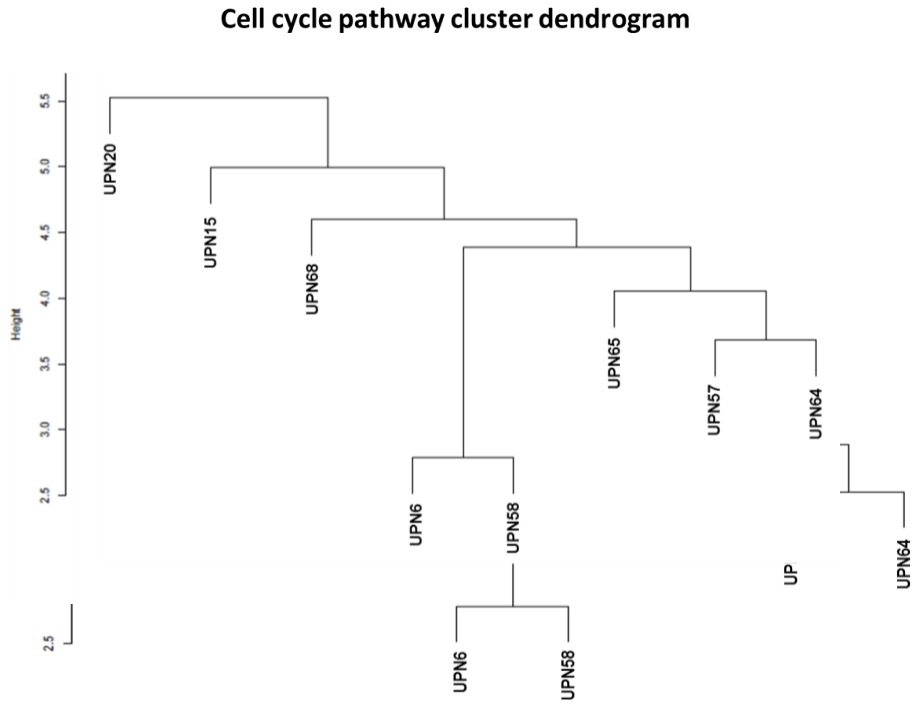

Fig 1b: Focus on the dendrogram showing the clusterization of SDS patients according to the SNVs found.