From whole exome analysis in idiopathic azoospermia to the identification of a high risk subgroup for occult Fanconi Anemia

Csilla Krausz, MD, PhD[1,2†*], Antoni Riera-Escamilla, PhD[2†], Chiara Chianese, PhD[2], Daniel Moreno-Mendoza, MD[2], Elisabet Ars, PhD[2], Osvaldo Rajmil, MD, PhD[2], Roser Pujol, PhD[3], Massimo Bogliolo, PhD[3], Ignacio Blanco, MD, PhD[4], Ines Rodríguez, MD[4], Isabel Badell, MD, PhD[5], Eduard Ruiz-Castañé, MD[2], Jordi Surrallés, PhD[3*]

[1]Department of Experimental and Clinical Biomedical Sciences "Mario Serio", Centre of Excellence *DeNothe*, University of Florence, Florence, Italy

[2]Andrology Department, Fundació Puigvert, Universitat Autònoma de Barcelona, Instituto de Investigaciones Biomédicas Sant Pau (IIB-Sant Pau), Barcelona, Catalonia, Spain

[3]Genetics Department and Biomedical Reseach Institute, Hospital de Sant Pau, Center for Biomedical Research on Rare Diseases (CIBERER), and Department of Genetics and Microbiology, Universitat Autònoma de Barcelona, Barcelona, Catalonia, Spain

[4]Hospital Germans Trias i Pujol, Badalona, Catalonia, Spain

[5]Pediatrics Department, Hospital de Sant Pau, Barcelona, Catalonia, Spain

† Prof. Csilla Krausz and Dr. Riera-Escamilla contributed equally to this article

*Correspondece: Prof. Csilla Krausz (csilla.krausz@ unifi.it) and Prof. Jordi Surrallés (jsurralles@santpau.cat)

Keywords: whole-exome sequencing, genetics, male infertility, Fanconi Anemia, azoospermia

ABSTRACT

Purpose: In about 10% of patients affected by Fanconi Anemia (FA) the diagnosis is delayed until adulthood and the presenting symptom in these "occult" FA is often a solid cancer and cancer-treatment related toxicity. Highly predictive clinical parameter(s) for diagnosing such an adult onset cases are missing.

Methods: i) Whole-Exome Sequencing (WES) ii) Sanger sequencing of the *FANCA* gene; iii) DEB-induced chromosome breakage test.

Results: WES identified a pathogenic homozygous *FANCA* variant in a patient affected by Sertoli Cell Only Syndrome (SCOS) and in his azoospermic brother. Although they had no overt anemia, chromosomal breakage test revealed a reverse somatic mosaicism in the former and a typical FA picture in the latter. In 27 selected SCOS cases one additional patient showing compound heterozygous pathogenic *FANCA* variants was identified with positive chromosomal breakage test.

Conclusion: We report an extraordinary high frequency of FA in a specific subgroup of azoospermic patients (7.1%). The screening for *FANCA* mutations in such patients has the potential to identify undiagnosed FA before the appearance of other severe clinical manifestations of the disease. The definition of this high-risk group for "occult" FA, based on specific testis phenotype with mild/borderline hematological alterations, is of unforeseen clinical relevance.

INTRODUCTION

The severest form of male factor infertility is non-obstructive azoospermia (NOA), which occurs in approximately 1% of all men in reproductive age and in the majority of patients the etiology remains unknown.[1] Since spermatogenesis is regulated by the concerted action of >2000 genes, a large proportion of idiopathic NOA might be attributable to a not yet identified genetic defect.[2] Within the last years, Whole Exome Sequencing (WES) has proved to be a powerful tool for the discovery of causative variants in familial cases of NOA.[3–10] On the other hand, this approach has the potential to incidentally reveal undiagnosed diseases. Genes involved in stem cell proliferation and DNA repair are important not only for spermatogenesis but also for other physiological processes. For instance, the association between FA, caused by mutations in genes involved in DNA interstrand cross-link repair, and infertility is well known in humans; however, the exact testicular phenotype in FA patients has not been explored. FA arises in the large majority of cases with bone marrow failure during childhood however, in 10% of cases the diagnosis is delayed until adulthood.[11] The late diagnosis frequently occurs in those patients who present a slow progressive depletion of bone marrow hence the sign of onset may be the appearance of a cancer instead of an overt anemia. Diagnosing occult FA has a relevance also concerning cancer treatment due to acute toxicity to chemotherapy. We performed WES in an idiopathic NOA (iNOA) patient with consanguineous parents, who displayed Sertoli Cell-Only Syndrome (SCOS). Our investigation revealed a pathogenic variant in the *FANCA* gene providing an incidental diagnosis of FA during a genetic investigation for NOA in an adult man. Based on this finding we revisited all our patients with similar testis histology and selected among them those with mild alterations or borderline values of platelets and Mean Corpuscular Volume (MCV) for further screening of *FANCA* gene mutations. This approach has led to the identification of a high-risk group for adult onset FA.

MATERIALS AND METHODS

Study populations

Study population I: A 43-year-old man (04-170) with iNOA and with consanguineous parents (first-cousin marriage) was selected for WES. All known causes of azoospermia were excluded. Testis histology revealed SCOS type I (no germ cells were present in the tubules). See Table 1.

Study population II: 27 iNOA patients with testis histology of SCOS type I and SCOS type II (SCOS with germ cells in a minority of tubules), with platelets <200.000/L and MCV>85fL were selected for *FANCA* mutation screening (Table S1).

Whole Exome Sequencing (WES) in patient 04-170

Library preparation, Bioinformatic analysis and variants filtering are described in Supplementary Material.

Multiplex Ligation-Dependent Probe Amplification (MLPA)

Large deletions were analyzed using a Multiplex Ligation-Dependent Probe Amplification (MLPA) kit developed by MRC-Holland BV as we described before.[12]

FANCA Sanger sequencing analysis

Two-steps Sanger Sequencing (described by Castella et al. 2011)[12] in 27 selected iNOA patients: first, the three most commonly mutated exons (exons 13, 36, 38) in the Spanish population; second, the remaining *FANCA* exons.

Diepoxibutane (DEB)-induced chromosomal breakage test: mutation carrier NOA patients and and their brother's peripheral blood lymphocytes, as described elsewhere[13] (see also Supplementary Material).

Gene Expression Evaluation

Quantitative RT-PCR (qRT-PCR) analysis was performed to evaluate *FANCA* expression in adult testes biopsies, collected in our clinic and characterized with molecular markers for the definition of distinct germ cell types (details in Supplementary Material and Table S2).

RESULTS

Exome analysis of a NOA patient affected by SCOS

Patient 04-170 carried a total of 20477 exonic variants. Following standard filtering (described in supplementary material) and the exclusion of synonymous variants with a MAF≥0.05, we filtered for all homo/hemizygous variants that presented an in-house Index of Pathogenicity (IP) ≥0.7. Of the eight resulting variants, only one mapped to a gene with proved implication in spermatogenesis (*FANCA)* and presented the highest

pathogenic score (IP=0.9,) (Figure 1, Figure S1 and Table S3). The c.2639G>A variant maps to exon 28 of the *FANCA* gene (NM_000135.2) and its pathogenicity was proven by functional studies[12,14] . It was already shown to cause FA in a Spanish patient,[12] as well as reported in the FA Mutation Database at Rockefeller University both in homozygosis and heterozygosis in eight FA patients (Table S4). DEB-induced chromosomal breakage test, one of the gold-standard tests used in the routine practice for the diagnosis of FA, gave a positive result in a fraction of cells, indicating FA somatic mosaicism (Figure 1B). The presence of 2-3% of wild-type alleles in the NGS data (data not shown) suggests a possible mechanism of genetic revertion by back mutation. The confirmation of this hypothesis would request additional experiments using digital PCR, cloning and single clone sequencing or amplicon analysis by deep NGS[15].

Following the discovery of the variant, the patient, during genetic counseling, informed the doctor that also his brother suffered by azoospermia. The screening of *FANCA* in his brother revealed the same genotype. The man could not undergo physical examination in our clinic due to geographic distances, but according to the family doctor he has no evident skeletal abnormalities or other FA-related symptoms. DEB-induced chromosomal breakage test showed that the majority of cells resulted positive, revealing a typical complete FA picture (Figure 1B). It is interesting to note that although this man presented a mild decrease of platelets in 2011 and, as for 2013 he also displayed a mild decrease of red blood cells and leucocytes, this condition was not further explored by the family doctor. The hemogram from 2016, requested after our diagnosis, shows a pronounced decrease of all three cell types (Table 1).

FANCA mutation screening in selected iNOA patients with SCOS

From our large cohort of 1300 infertile men, 136 were classified as iNOA with testis histology of SCOS type I or type II. For 120/136 patients hematological parameters were available and we selected those (n=27) with platelets <200.000/L and Mean Corpuscular Volume >85fL (Figure S2, Table S1). Genomic DNA from peripheral blood of this specific group of 27 iNOA patients was screened for mutation firstly in exons 13, 36, 38 i.e. those exons which contain the most commonly observed mutations in Spanish FA patients and for large *FANCA* deletions through MLPA. We identified a heterozygous variant in exon 38 in patient 14-339 [NM_000135.2:c.3788_3790delTCT, p.Phe1263del)]. This analysis was followed by Sanger sequencing of the entire *FANCA* gene which revealed a second heterozygous variant in exon 39

[NM_000135.2:c.3913C>T ,p.Leu1305Phe] in patient 14-339. Both mutations were previously reported in many FA patients and are clearly pathogenic (Figure S2, Table S4). Similar to patient 04-170, DEB-induced chromosomal breakage test gave a positive result in a fraction of cells confirming mosaic FA (Figure S3). The brother of this patient showed normozoospermia and his hormone levels, testis volume and blood count were within the normal range (Table 1). Sanger Sequencing revealed only one of the two variants in heterozygosis [NM_000135.2:c.3913C>T, p.Leu1305Phe].

FANCA expression in the testis*:* qRT-PCR analysis was performed in a collection of testis biopsy samples (not belonging to the patients' testis biopsies due to the lack of material recollection) previously characterized by a molecular genetic approach (see Supplementary Material). *FANCA* was undetectable in SCOS, but showed the highest expression in biopsies with a prevalent spermatogonial arrest (SGA) and obstructive azoospermia (i.e. presence of all spermatogenesis stages). This expression pattern is compatible with a prevalent expression in spermatogonial and early spermatocytic stage (Figure S4). .

DISCUSSION

The large majority of FA patients are diagnosed in childhood due to the presence of significant birth defects. About 10% of patients the diagnosis is delayed until adulthood,[11] especially when they appear normal or have subtle findings that may be overlooked. The late diagnosis frequently occurs in those patients who present a slow progressive depletion of bone marrow hence the sign of onset may be the appearance of a cancer instead of an overt anemia. According to the FA Guidelines,[16] "male infertility" is one of those adult-onset signs/symptoms which should raise suspicion of FA, hence should prompt clinicians to perform peripheral blood testing. However, while FA is a rare disease (prevalence of 1/160.000), infertility is a common pathological condition, present in 7% of the general male population. Moreover, it is an extremely heterogeneous condition ranging from the complete absence of spermatozoa in the ejaculate (azoospermia) to decreased sperm number or motility/morphology.[1] Even the NOA phenotype can be further sub-classified according to the testis histology e.g. complete absence of germ cells (SCOS), spermatogenic arrest (at different stages) or hypospermatogenesis. Therefore, referring to "infertility" as a possible FA sign is extremely broad and requires a more precise definition in the clinical practice.

Our first patient, diagnosed through WES, presented borderline values of blood counts and besides infertility showed another potential FA sign, mild dysmorphic facies. His brother with the same reproductive problem carried the same genotype and as expected, the chromosomal breakage test confirmed FA in both brothers. A picture of mosaic FA in patient 04-170 and a complete FA in his brother who did not show dysmorphic features but displayed a progressive impairment of hematopoiesis reaching to pathologically low values at the control in 2016 (performed after our diagnosis).
The third patient presented low stature and some *cafe au lait* spots in his hands and back; however, FA was not suspected until he came to our clinic and was diagnosed with azoospermia. Similar to patient 04-170, DEB test showed a picture of mosaic FA also in this man. The Patient 14-339's brother carried only one of the variants in heterozygosis and his testicular function was normal. A common feature of all three affected men were their testis phenotype i.e. SCOS.
In this study, we report an unexpectedly high frequency of FA (2/28=7.1% or 3/29 = 10.3% if we consider also the brother) in a specific subgroup of iNOA subjects with testicular phenotype of SCOS type I or type II and with slightly altered/borderline hematological parameters. Our study indicates that andrological evaluation, especially in SCOS patients, should not only include hormone measurement but also blood count, since it may lead to the diagnosis of unsuspected FA in adulthood. It further supports that FA is not an exclusive childhood illness, and diagnosis and treatment are no longer exclusively performed by pediatricians. We propose that the evaluation of hematological parameters, in particular platelet count and MCV in NOA patients should be included in the routine andrological workup. *Vice versa*, we provide a more precise andrological criteria to all clinicians, about the type of "infertility" subgroup in which the likelihood of FA is relatively high. The combined assessment of testis histology and hematological parameters allows the diagnosis of "occult" FA before the appearance of severe pathologies related to this disease (e.g. hematological malignancies and specific solid cancers). It has been reported that in more than 20% of patients with FA who developed solid tumors, the diagnosis of FA was made only after the appearance of their cancer and its treatment.[17] (some genotoxic anticancer agents are contraindicated in FA patients because of their DNA repair deficiency).The diagnosis of *FANCA* variants in these three men has also relevance in case they develop this type of cancers in the future and thanks to our investigation, they are now receiving specific medical attention including strict follow-up by oncohematologists. The link between severe

spermatogenic impairment and chronic diseases has been advocated by epidemiological observations, showing that infertile men (oligo/azoospermic) have a higher risk of morbidity (including cancer) and a lower life expectancy.[18–20] A biological basis for such observations may be: i) the previously observed high CNV burden[2] (indicating chromosomal instability) in infertile men; ii) that some genes involved in spermatogenesis may also have a role in carcinogenesis (especially those involved in stem cell renewal/differentiation and mismatch repair mechanisms[21]). Besides its clinical importance, our study stimulates further investigations on the discovery of common genetic factors involved both in SCOS and bone marrow failure.

Acknowledgements: This work was funded by the Spanish Ministry of Health-Instituto Carlos III-FIS (grant number: FIS/FEDER-PI14/01250; PI17/01822) and European Commission, Reproductive Biology Early Research Training (REPROTRAIN, Project Number: 289880) awarded to CK and AR-E. JS is funded by ICREA Academia, EUROFANCOLEN project HEALTH-F5-2012-305421, CIBERER and SAF2015-64152-R/FEDER. The authors wish to thank the patients participating in the study for their important collaboration. We also thank the clinicians from the Andrology Unit of the Fundació Puigvert J.Sarquella, J.Sanchez-Curbelo. We acknowledge Esperança Martí, President of the Fundació Puigvert.

Disclosure:

The authors have nothing to disclose.

Table 1. Clinical characteristics of the patients carrying *FANCA* variants and of their brothers

	04-170	04-170 brother	14-339	14-339 brother
Andrological/ Endocrine and metabolic parameters				
Semen phenotype/ Testis histology	Azoospermia/ SCOS type I	Azoospermia/ not performed	Azoospermia/ SCOS type II	Normozoo-spermia
Age at first semen analysis/Testis biopsy	32	n.a	34	33
Testis volume, left;right (ml)	5; 12	n.a.	12; 10	25; 25
Relevant clinical findings	Bilateral cryptorchdism; mild dismorphic *facies*	n.a	Low stature, *cafe au lait* spots	No
FSH (IU/L)	25.6	23.3	21.3	4.54
LH (IU/L)	5.9	10.2	7.53	4.9
T (nmol/L)	15.3	8.2	20.2	18.8
Height (cm)	175	178	158	179
BMI (kg/m^2)	26.12	26.51	20.83	20.29
Glycaemia (mmol/L)	5.7	4.61	5.3	5.0
Cholesterol (mmol/L)	3.47	5.19	4.2	4.61
Hematological parameters				
Erythrocyte(10^{12}/L) *Reference Values: 4.30-6.00*	4.94	4.27	3.43	4.86
Hemoglobin (g/L) *Reference Values: 140-180*	161	136	132	145
Hematocrit (L/L) *Reference Values: 0.40-0.52*	0.48	0.40	0.38	0.43
Leucocytes (10^9/L) *Reference Values: 4.00-10.50*	4.02	3.02	3.70	5.39
Platelets (10^9/L) *Reference Values: 150-400*	194	116	137	174
MCV (fL) *Reference Values: 80-100*	98.10	92.97	112.30	89.30
MCH (pg) *Reference Value 27-34*	32.60	31.85	38.60	29.80
Neutrophils *Reference Values: 50-70*	49.30%	48.00%	49.60%	43.50%
Lymphocytes *Reference Values: 16-45*	36.10%	41.10%	32.90%	38.60%
Monocytes *Reference Values: 3-12*	9.60%	9.30%	15.90%	8.20%
Eosinophils *Reference Values: 0.40-6.60*	1.10%	1.30%	1.10%	9.10%
Basophils *Reference Values < 1.00*	0.40%	0.30%	0.50%	0.60%

SCOS type I: Sertoli Cell Only Syndrome without germ cells in **all** tubules; SCOS type II: Sertoli Cell Only Syndrome with germ cells in a minority of tubules. n.a. not available

REFERENCES:

1 Tournaye H, Krausz C, Oates RD. Novel concepts in the aetiology of male reproductive impairment. *lancet Diabetes Endocrinol* 2016; published online July. DOI:10.1016/S2213-8587(16)30040-7.

2 Krausz C, Escamilla AR, Chianese C. Genetics of male infertility: from research to clinic. *Reproduction* 2015; **150**: R159–74.

3 Ayhan Ö, Balkan M, Guven A, *et al.* Truncating mutations in TAF4B and ZMYND15 causing recessive azoospermia. *J Med Genet* 2014; **51**: 239–44.

4 Okutman O, Muller J, Baert Y, *et al.* Exome sequencing reveals a nonsense mutation in TEX15 causing spermatogenic failure in a Turkish family. *Hum Mol Genet* 2015; **24**: 5581–8.

5 Ramasamy R, Bakırcıoğlu ME, Cengiz C, *et al.* Whole-exome sequencing identifies novel homozygous mutation in NPAS2 in family with nonobstructive azoospermia. *Fertil Steril* 2015; **104**: 286–91.

6 Gershoni M, Hauser R, Yogev L, *et al.* A familial study of azoospermic men identifies three novel causative mutations in three new human azoospermia genes. *Genet Med* 2017; published online Feb. DOI:10.1038/gim.2016.225.

7 Kherraf Z-E, Christou-Kent M, Karaouzene T, *et al.* SPINK2 deficiency causes infertility by inducing sperm defects in heterozygotes and azoospermia in homozygotes. *EMBO Mol Med* 2017; **9**: 1132–49.

8 Arafat M, Har-Vardi I, Harlev A, *et al.* Mutation in TDRD9 causes non-obstructive azoospermia in infertile men. *J Med Genet* 2017; published online May. DOI:10.1136/jmedgenet-2017-104514.

9 Okutman O, Muller J, Skory V, *et al.* A no-stop mutation in MAGEB4 is a possible cause of rare X-linked azoospermia and oligozoospermia in a consanguineous Turkish family. *J Assist Reprod Genet* 2017; **34**: 683–94.

10 Colombo R, Pontoglio A, Bini M. Two Novel TEX15 Mutations in a Family with Nonobstructive Azoospermia. *Gynecol Obstet Invest* 2017; **82**: 283–6.

11 Alter BP. Bone marrow failure: a child is not just a small adult (but an adult can have a childhood disease). *Hematol Am Soc Hematol Educ Progr* 2005; : 96–103.

12 Castella M, Pujol R, Callen E, *et al.* Origin, functional role, and clinical impact of Fanconi anemia FANCA mutations. *Blood* 2011; **117**: 3759–69.

13 Castella M, Pujol R, Callen E, *et al.* Chromosome fragility in patients with Fanconi anaemia: diagnostic implications and clinical impact. *J Med Genet* 2011; **48**: 242–50.

14 Mankad A, Taniguchi T, Cox B, *et al.* Natural gene therapy in monozygotic twins with Fanconi anemia. *Blood* 2006; **107**: 3084–90.

15 Asur RS, Kimble DC, Lach FP, *et al.* Somatic mosaicism of an intragenic FANCB duplication in both fibroblast and peripheral blood cells observed in a Fanconi anemia patient leads to milder phenotype. *Mol Genet genomic Med* 2018; **6**: 77–91.

16 Fanconi Anemia Research Fund I. Fanconi Anemia : Guidelines for Diagnosis and Management (Fourth edition). 2014.

17 Alter BP, Joenje H, Oostra AB, Pals G. Fanconi anemia: adult head and neck cancer and hematopoietic mosaicism. *Arch Otolaryngol Head Neck Surg* 2005; **131**: 635–9.

18 Jensen TK, Jacobsen R, Christensen K, Nielsen NC, Bostofte E. Good semen quality and life expectancy: a cohort study of 43,277 men. *Am J Epidemiol* 2009; **170**: 559–65.

19 Eisenberg ML, Li S, Brooks JD, Cullen MR, Baker LC. Increased risk of cancer in infertile men: analysis of U.S. claims data. *J Urol* 2015; **193**: 1596–601.

20 Latif T, Kold Jensen T, Mehlsen J, *et al.* Semen quality is a predictor of subsequent morbidity. A Danish cohort study of 4,712 men with long-term follow-up. *Am J Epidemiol* 2017; published online May. DOI:10.1093/aje/kwx067.

21 Maduro MR, Casella R, Kim E, *et al.* Microsatellite instability and defects in mismatch repair proteins: a new aetiology for Sertoli cell-only syndrome. *Mol Hum Reprod* 2003; **9**: 61–8.

FIGURE LEGEND

Figure 1. Investigation on the patient 04-170 and his brother, both carriers of the FANCA variant. A) The inverted pyramid scheme indicates how WES data were filtered and prioritized for patient 04-170 starting from a total of 20477 variants. Eight homozygous, rare and potentially pathogenic variants were identified in eight different genes (in bold gene expressed in testis). B) Graphic representation of the chromosomal breakage test for patient 04-170 and his brother. The arrows illustrate the patient's and his brother's position in relation to the distribution of historical data collected in our laboratory according to the percentage of DEB-induced aberrant cells and the number of breakages in each DEB-induced aberrant cell. In the No FA group only individuals with at least one aberrant cell were included. C) The pedigree structure shows the segregation of the c.2639C>T (p.Arg880Gln) variant. Colored symbols are explained in the legend in the lower part of the figure. The lower-left arrow indicates the proband 04-170 who was subjected to WES analysis whereas his brother' DNA was sequenced by Sanger sequencing .. Our patient and his brother both carried the c.2639G>A (p.Arg880Gln) variant in homozygosis and were azoospermic. The parents' DNA samples were not available for testing, but being the variant rare it can be assumed that the parents were both heterozygous carriers

UNIVERSITÀ DEGLI STUDI FIRENZE

DIPARTIMENTO DI SCIENZE BIOMEDICHE SPERIMENTALI E CLINICHE
VIALE MORGAGNI, 50
FIRENZE

Florence, 21st November 2017

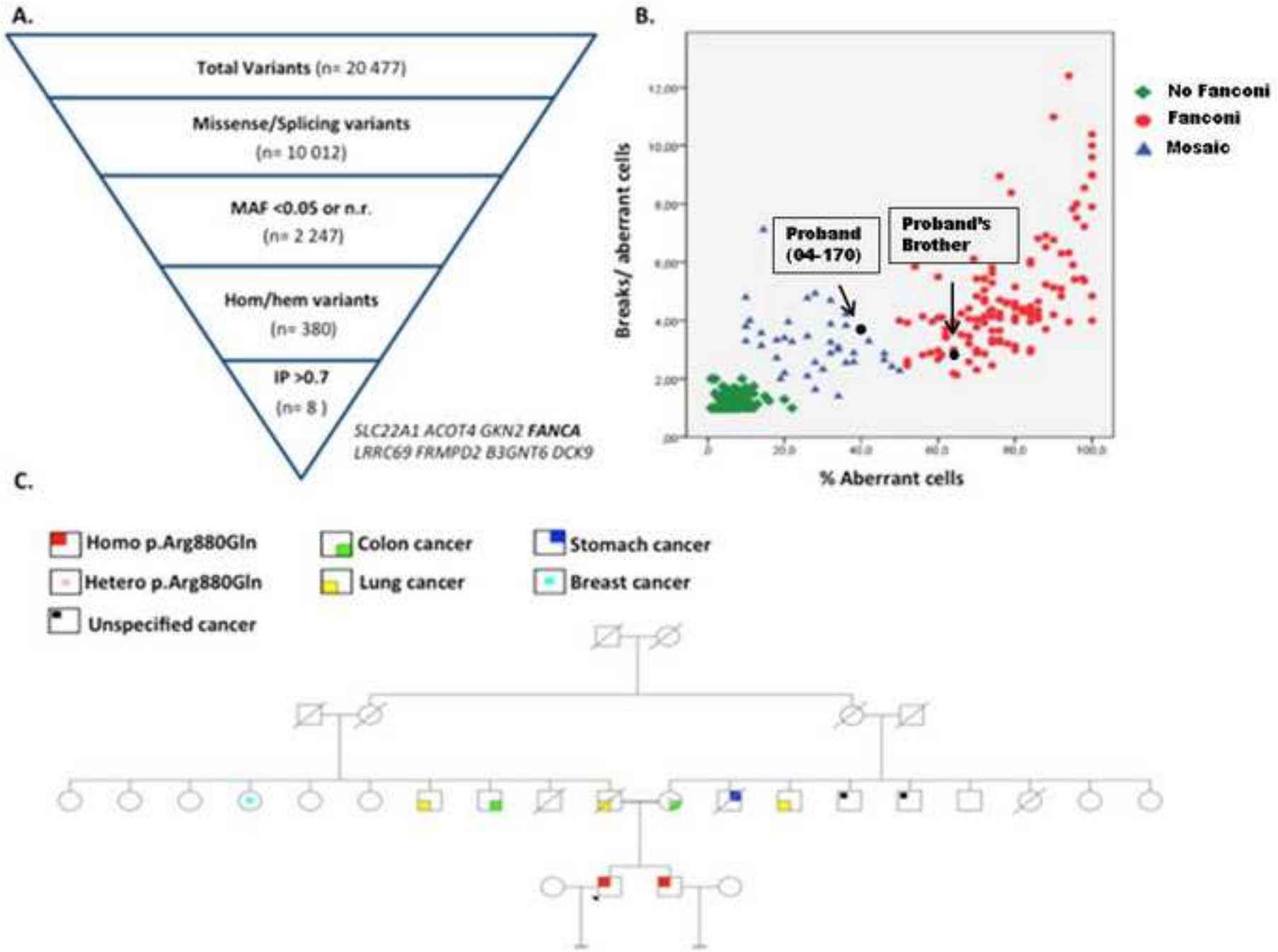
Conflict of interest declaration:

Concenring the manuscript entitled "**From whole exome analysis in idiopathic azoospermia to the identification of a high risk subgroup for occult Fanconi Anemia"** to be considered for publication in Genetics in Medicine we declare no conflict of interest. No Pharmaceutical industry has been involved. The work in this paper has been funded by the Spanish Ministry of Health FIS awarded to Csilla Krausz (grant number: FIS/FEDER PI14/01250) and Jordi Surrallés is funded by ICREA Academia, EUROFANCOLEN project HEALTH-F5-2012-305421, CIBERER (CIBERER is an initiative of the Instituto de Investigación Carlos III) and SAF2015-64152-R/FEDER

On behalf of the authors,

Csilla Krausz MD, PhD
Professor in Endocrinology

Figure 1



SUPPLEMENTARY INFORMATION

Supplement to Krausz C, Riera-Escamilla A, Chianese C et al. , **From whole exome analysis in idiopathic azoospermia to the identification of a high risk subgroup for occult Fanconi Anemia**

Contents:

Author List

Csilla Krausz[1,2]†* M.D., Ph.D., Antoni Riera-Escamilla[2]† Ph.D, Chiara Chianese[2] Ph.D., Daniel Moreno-Mendoza[2] M.D., Elisabet Ars[2] Ph.D, Osvaldo Rajmil[2] M.D., Ph.D., Roser Pujol[3] Ph.D, Massimo Bogliolo[3] Ph.D., Ignacio Blanco[4] M.D., Ph.D., Ines Rodríguez[4] M.D., Ph.D., Isabel Badell, M.D., PhD[5] Eduard Ruiz-Castañé[2] M.D., Jordi Surrallés[3*] Ph.D.

† Equally contributing authors

1. Department of Experimental and Clinical Biomedical Sciences "Mario Serio", Centre of Excellence *DeNothe*, University of Florence, Italy

2. Andrology Department, Fundació Puigvert, Universitat Autònoma de Barcelona, Instituto de Investigaciones Biomédicas Sant Pau (IIB-Sant Pau), Barcelona, Catalonia, Spain

3. Genetics Department and Biomedical Reseach Institute, Hospital de Sant Pau, Center for Biomedical Research on Rare Diseases (CIBERER), and Department of Genetics and Microbiology, Universitat Autònoma de Barcelona, Barcelona, Catalonia, Spain.

4. Hospital Germans Trias i Pujol, Badalona, Catalonia, Spain.

5. Pediatrics Department, Hospital de Sant Pau, Barcelona, Catalonia, Spain

[*]Email corresponding authors: c.krausz@dfc.unifi.it; jsurralles@santpau.cat.

Supplementary Materials and Methods

Whole-exome Sequencing

The analysis (starting from the standard DNA quality check) has been provided as SERVICE by BGI (BGI TECH SOLUTIONS (Hong Kong). The description below derives from the limited information provided by the company.

After the standard DNA quality control steps, whole- exome sequencing capture was performed using the BGI Human 59M Exon kit based on Combinatorial Probe- Anchor Ligation (cPAL™) technology[1] and the captured material was sequenced on Complete Genomics platform as service. According to the information provided by the service provider, the qualified genomic DNA was randomly fragmented to an average size of 200-400 bp and then, AdA 5'- and 3'-adaptors were ligated to the 5'- and 3'-ends of the fragments, respectively. The AdA adaptor-ligated fragments were amplified by PCR, and the PCR products were used for the follow-up exon captured (BGI exome V4, 59Mb).

During the exon capture, the exon fragments hybridized with the capture RNA 100nt kit probes stably and were therefore captured effectively. The captured exon fragments were purified by DynabeadsM-280 Streptavidin Bead purification and were further amplified by another round PCR. Then, the PCR products were circularized and the resulting double strand (ds) circles were digested with Ecop15. Among these digested fragments, small fragments were collected after bead purification. Similar to the AdA adaptor ligation, AdB were ligated to both end of the purified fragments (aforementioned). These fragments were denatured into two single strands (ss) and the target strands were selected for the circularization. The ss circles as templates were amplified to be DNA nanoballs (DNBs). DNBs were loaded on the slides and sequenced on the Complete Genomics' platform with a 100x mean coverage depth.

Bioinformatic analysis

Sequence reads were mapped to the reference genome (GRCh37) and variants were called in the aligned genome according to the methods described by Carnevali et al 2012[2] using version 2.4 of the Complete Genomics software. Briefly, exome reads were analyzed using a proprietary technology based on Teramap for alignment on GRCh37

(hg19) reference sequence. Variants with low-quality reads were excluded to guarantee ≥ 80% of bases with quality score of ≥Q20. According to the alignment results, regions of the genome deemed likely to differ from the reference genome were identified. Then, individual reads that were likely to lie in those regions were collected and a local de novo assembly was performed. Next, based on the initial mapping and assembly results, a probability statistical model (Bayesian Modeling) was applied to call variants by computing a probability ratio for any two hypotheses from the optimization step; variant calls are then made based on the most likely hypothesis according to this Bayesian probability model. Variants extracted from those hypotheses with a likelihood exceeding the significance threshold were reported.

Variants filtering

Firstly, standard filtering was applied to all samples. Briefly, we selected missense variants, stop gains/losses, frameshift insertions/deletions, and filtered out common polymorphisms (≥5% in the general population) after consulting the dbSNP 138 and the 1000G (http://www.1000genomes.org). After applying the abovementioned filter, we filtered for hemizygous and homozygous variants for which the global minor allele frequency (MAF) was ≤0.05 or not reported. Overall, obtained data was further filtered according to their potentially damaging effect depending on the type of variant. For SNVs, the prediction of pathogenicity was based on SIFT, Polyphen2, Mutation Taster, Mutation Assessor, RadialSVM, LRT and LR prediction tools. An index of pathogenicity was created as a score calculated on the basis of the seven prediction tools employed, each providing a value ranging from -1 (null probability of being a deleterious variant) to 1 (full probability of being a deleterious variant), as illustrated in Table S3. Not all prediction tools were always available; therefore a ratio was calculated between the summary score of pathogenicity and the number of prediction tools available for a determined variant (see table S3) We filtered in variants with IP>0.7 which corresponds to an arbitrary threshold, which means that the majority (not necessary all) of the interrogated tools gave a "pathogenic" score. We considered of further interest only filtered homozygous variants that were present in genes with potential implication in the early phases of spermatogenesis (Figure S1A). Concerning indels and nonsense variants prediction tools are not always available, a different approach was employed for the selection of candidate variants: all patient-specific homo/hemizygous variants with MAF≤0.01 were filtered in. For all variants, SNVs and

indels, a thorough investigation of the literature (including papers which are dealing with genes involved in azoospermia in consanguineous families) and on-line databases was performed to select variants in genes that had a potential role in the early phases of spermatogenesis. Genes for which no data on fertility were available as well as genes with no putative implication in the early stages of spermatogenesis were excluded (Figure S1B). Then the Integrative Genomics Viewer (IGV) was employed to exclude eventual false positive calls and variants prioritized as described above were validated by Sanger sequencing (Figure S1A).

Molecular characterization of testis biopsy samples

RNA was isolated from snap frozen adult testis biopsies collected from azoospermic patients with different testis histology (obstructive azoospermia with normal spermatogenesis, SCOS, spermatogenic arrest at various stages). The molecular characterization consisted in the expression analysis of four genes known to be expressed in different stages of spermatogenesis: *DAZ* (spermatogonia/early spermatocytes), *CDY1* (Spermatids), *BRDT* (pachytene spermatocytes/round and elongating spermatids) and *PRM2* (spermatids/mature spermatozoa). The housekeeping reference gene was *GAPDH*. Based on the obtained expression profiles we selected three SCOS sample, one with spermatogenic arrest predominantly at spermatogonial stage (SGA), one sample with spermatocytic arrest (SCA) and two samples with obstructive azoospermia (OA) i.e. conserved spermatogenesis including mature spermatozoa. The extraction was performed through a combination of two commercially available kits, the TRI Reagent (Sigma-Aldrich, St. Louis, MO, USA) and the AllPrep DNA/RNA kit (Sigma-Aldrich, St. Louis, MO, USA), according to the manufacturer's instructions. cDNA synthesis was carried out using the High-Capacity cDNA Reverse Transcription Kit (Lifetechnologies, Foster City, CA, USA). qRT-PCR was performed using the TaqMan® Universal PCR Master Mix (Lifetechnologies, Foster City, CA, USA) with the following standard thermal cycler conditions: 40 cycles at 95 °C for 30 seconds and 60 °C for 1 min. Commercially available assays were employed to evaluate gene expression of four genes used for characterization(*BRDT, CDY1, DAZ, PRM2*) , the target gene (*FANCA*) and *GAPDH*, as reference gene for relative quantization of the target gene (Table S2).qRT-PCR runs were performed on a StepOne™ System (Applied biosystems, Carlsbad, CA, USA). Experiments were run in triplicates. Results concerning *FANCA* gene expression in the different testis biopsies are reported in Figure S4.

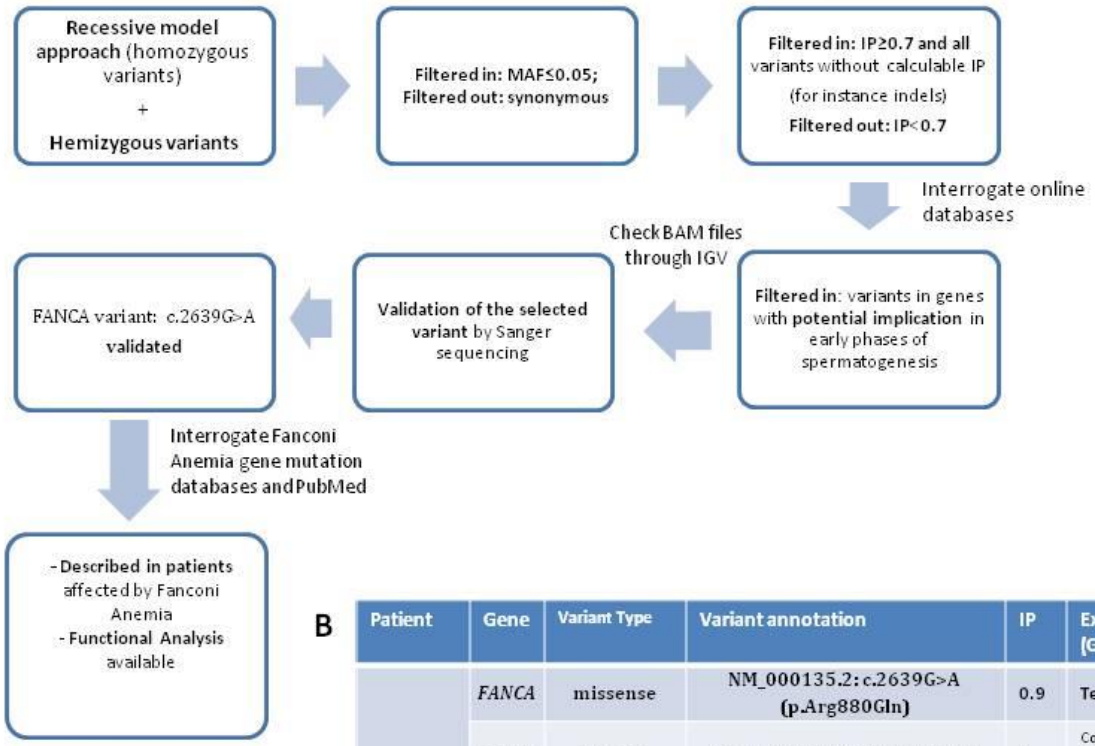
Detailed description of Diepoxibutane (DEB)-induced chromosomal breakage test

Three blood cultures were prepared for each patient, including 0.5 ml of blood in heparin and 4.5 ml of culture consisting of 15% fetal bovine serum, 1% antibiotics, 1% L-glutamine and 1% phytohaemagglutinin in RPMI (all reagents from Gibco, Carlsbad, CA, USA). Twenty four hours after culture set-up, two cultures were treated with DEB at a final concentration of 0.1 ug/ml (Sigma, Cat. No 202533, St. Louis, MO, USA), and the remaining culture was left untreated for spontaneous chromosome fragility evaluation. Forty-six hours after DEB treatment, colcemid was added at a final concentration of 0.1 ug/ml. Cultures were harvested 2 h later when metaphase spreads were obtained according to standard cytogenetic methods and finally, stained with Giemsa. For chromosome fragility evaluation, 25-50 metaphases with 46 centromeres were analyzed for each culture. The microscopic analysis was performed with a LeitzAristoplan microscope and, later with a Zeiss Imager M1 microscope coupled to a computerassisted metaphase finder (Metasystems, Altlussheim, Germany). The main criteria for the determination of chromosome fragility were as follows: gaps were not counted as chromosome breaks and figures were converted to the minimum number of breaks necessary to form each figure. DEB stock was routinely replaced every 6 months. Before using a new lot, a control fragility assay was performed using an FA lymphoblastoid cell line to ensure that there was no significant variation between lots.

<u>Figure S1</u>. Flowchart inherent to the exome analysis of patient 04-170.

The scheme briefly resumes how variants were filtered and then prioritized in patient 04-170, reaching to the identification and validation of the *FANCA* variant. A) Criteria used for filtering (after standard quality control) and prioritization; B) List of the filtered variants and their prioritization according to their expression in testis and their index of pathogenicity (IP). The only gene with testis expression is reported in bold (*FANCA*).

A



B

Patient	Gene	Variant Type	Variant annotation	IP	Expression in tissues (Genecards database)
04-170	***FANCA***	**missense**	**NM_000135.2: c.2639G>A (p.Arg880Gln)**	**0.9**	**Testis**
	B3GNT6	missense	NM_1387064: c.989T>G (p.Leu330Arg)	1	Colon-transverse, stomach, Small Intestine Terminal Ileum, and Esophagus - Mucosa
	DOCK9	missense	NM_015296.2: c.6098C>T (p.Thr2033Met)	1	Breast
	FRMPD2	splice_acceptor	NM_001018071.3:c.922G>A	1	Brain
	LRRC69	splice_donor	NM_001129890.1:c.934T>A	1	Bone marow, mesenchimal cells
	SLC22A1	missense	NM_003057.2: c.113G>A (p.Gly38Asp)	0.9	Liver
	GKN2	missense	NM_182536.2: c.193G>T (p.Asp65Tyr)	0.9	Stomach
	ACOT4	missense	NM_152331.3: c.568T>A (p.Tyr190Asn)	0.7	Liver, kidney, cortex

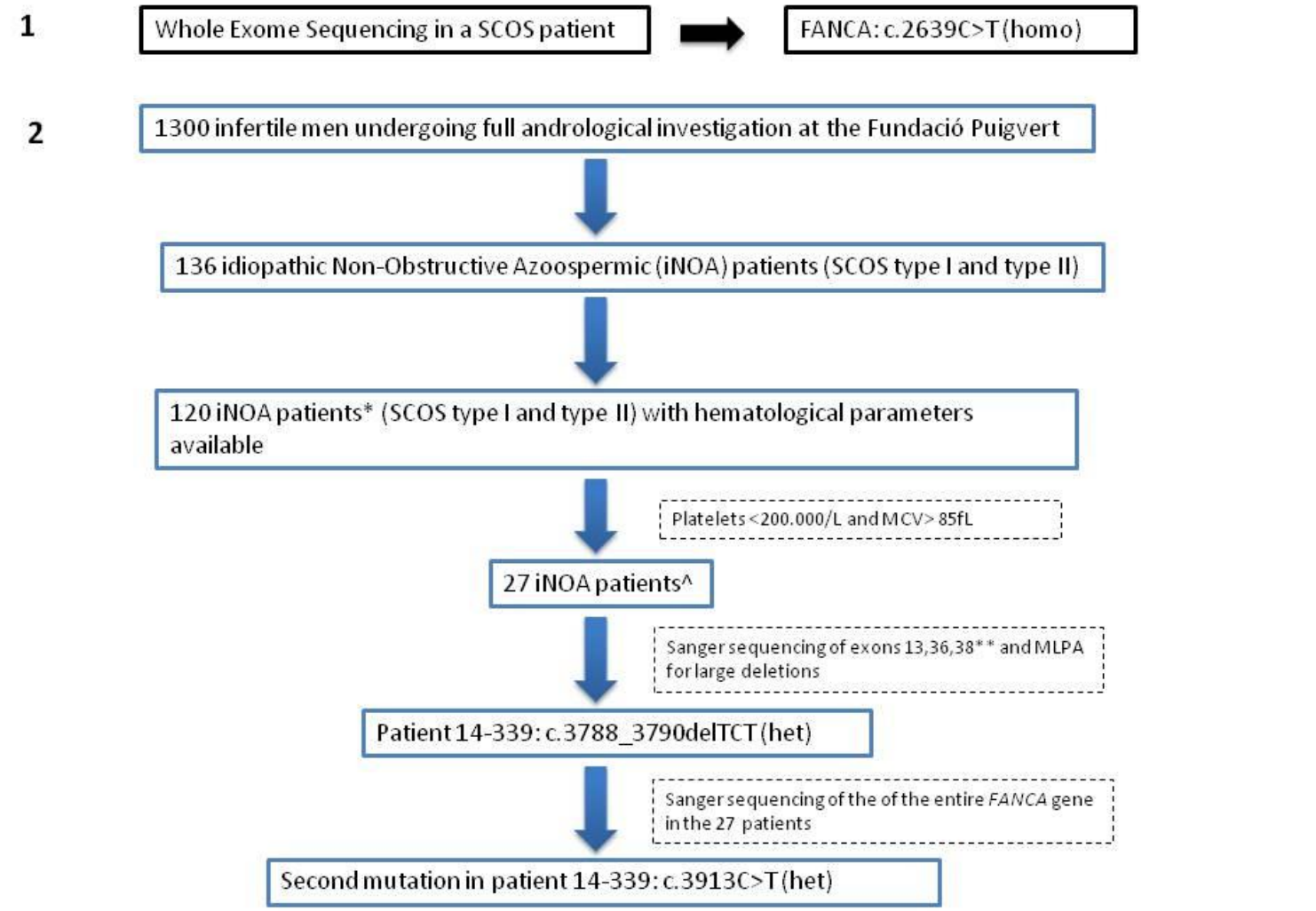
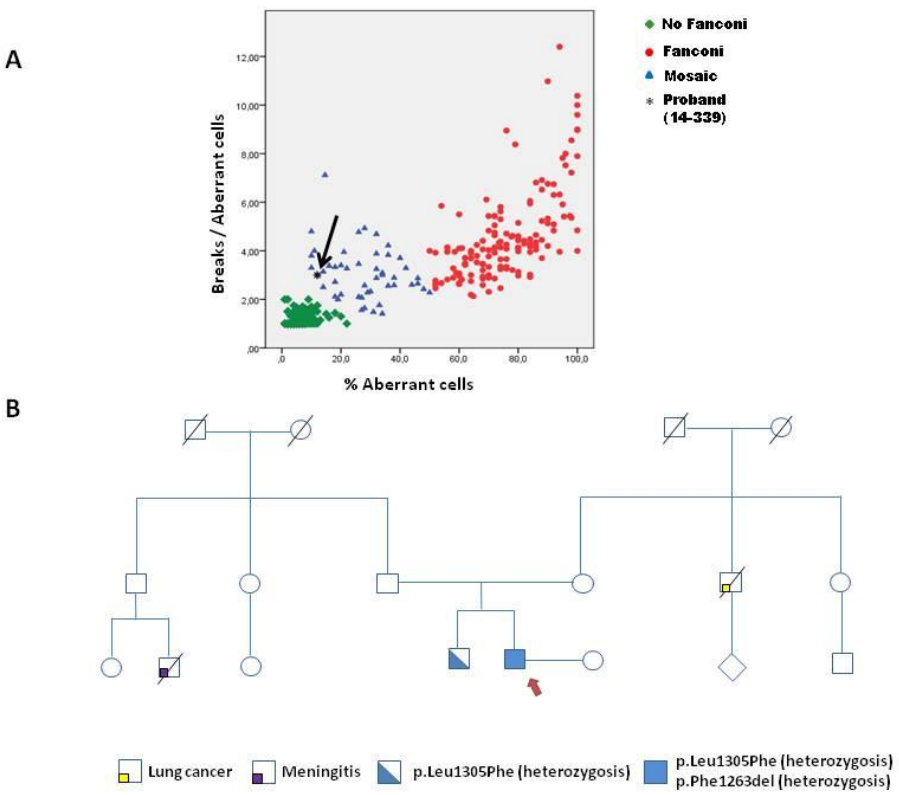


Figure S2. Flowchart reporting the study design leading to the identification of *FANCA* mutation carriers. A) The patient analyzed by Whole exome Sequencing and the mutation identified is reported. B) The scheme briefly resumes the criteria according to which infertile patients were selected and their DNA sequenced based on their testis histology, hematological parameters and genetic finding



<u>Figure S3.</u> Investigation on the patient 14-339 and his brother. Graphic representation of the chromosomal breakage test for 14-339. The graph illustrates the patient's position (*) in relation to the distribution of historical data collected in our laboratory. The pedigree structure shows the segregation of the p.Phe1263del and p.Leu1305Phe. Colored symbols are explained in the legend in the lower part of the figure. Patient 14-339 and his brother were sequenced for FANCA mutations. The lower-left arrow indicates the proband 14-339. The brother who was normozoospermic only carried the p.Leu1305Phe variant in heterozygosis. The parents' DNA samples were not available for testing, but being the variant rare it can be assumed that the parents were both heterozygous carriers

<u>Figure S4</u>. Expression evaluation of the *FANCA* gene

Quantitative RT-PCR (qRT-PCR) analysis was performed to evaluate *FANCA* expression in biopsy samples of different types of adult testis histologies: i) three SCOS (Sertoli Cell-Only Syndrome); ii) one SGA: maturation arrest at the spermatogonial level; iii) one SCA: maturation arrest at the spermatocytic level. Two samples with obstructive azoospermia (OA) were used as internal controls. Samples were first characterized by testing for four spermatogenic markers expressed at different stages of spermatogenesis: *PRM2* (spermatids/mature spermatozoa); *CDY1* (spermatids); *BRDT* (pachytene spermatocytes/round and elongating spermatids) and *DAZ* (spermatogonia/early spermatocytes)

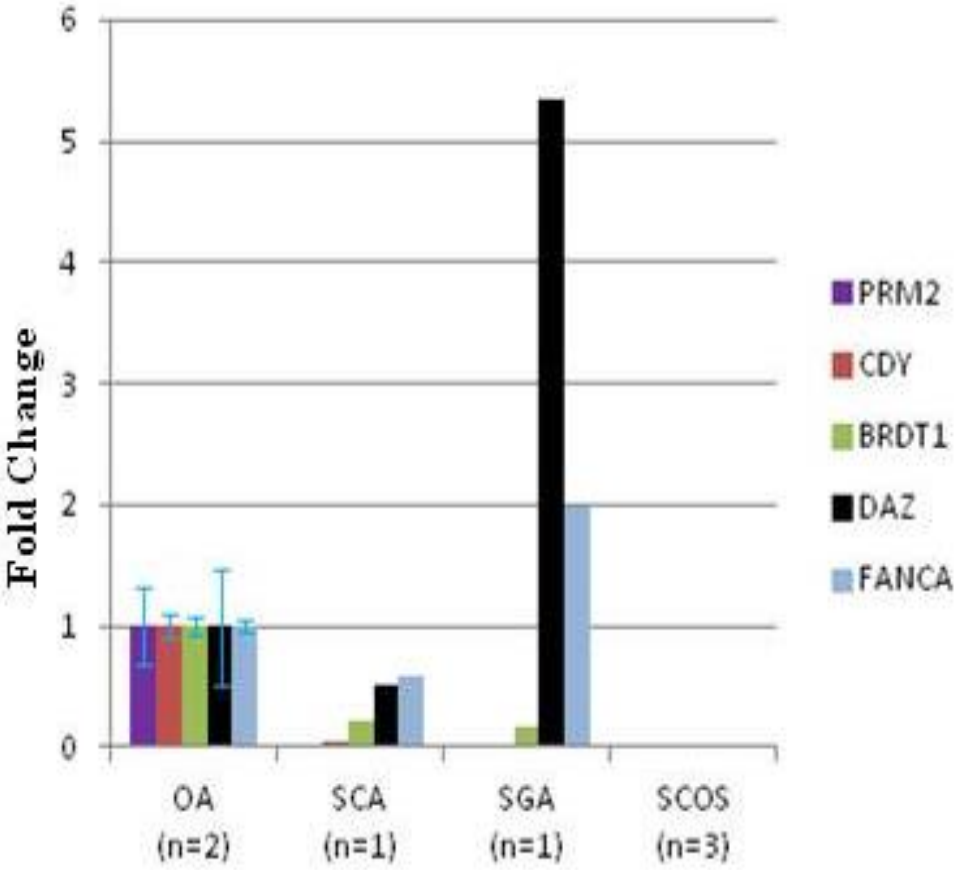


Table S1. Clinical characteristics of the 27 selected iNOA patients for *FANCA* gene mutational screening

Patient code	Semen Phenotype	Testicular phenotype	Platelets (10^9/L)	MCV (fL)	Erythrocytes (10^{12}/L)	Leucocytes (10^9/L)	Neutrophils (%)
14-339*	Azoospermia	SCOS type II	137	112.3	3.43	3.7	49.5
10-331	Azoospermia	SCOS type II	103	93.1	4.75	6	46.4
11-332	Azoospermia	SCOS type I	116	90.6	4.43	3.8	56.9
10-281	Azoospermia	SCOS type I	135	99.2	4.82	6.95	55.7
11-585	Azoospermia	SCOS type I	142	92.2	5.41	5.24	54.4
12-603	Azoospermia	SCOS type I	156	89.5	4.59	7.46	57.8
12-612	Azoospermia	SCOS type I	159	92.9	3.09	6.06	60.4
09-195	Azoospermia	SCOS type II	165	85.1	5.23	8.82	75.8
10-203	Azoospermia	SCOS type II	166	89.2	5.76	7.15	54.3
08-457	Azoospermia	SCOS type I	175	85.1	5.24	5.71	45.7
16-394	Azoospermia	SCOS type II	175	89.1	5.13	4.51	48.8
05-205	Azoospermia	SCOS type I	176	89.5	4.76	7.48	86.7
06-111	Azoospermia	SCOS type II	177	85.7	4.61	4.56	50.2
07-411	Azoospermia	SCOS type I	177	88.5	5.02	7.9	59.7
04-164	Azoospermia	SCOS type II	178	86.8	5.59	7.64	58.0
11-381	Azoospermia	SCOS type I	182	86.3	4.84	5.8	56.0
15-546	Azoospermia	SCOS type I	182	93.9	4.91	7.7	55.4
07-478	Azoospermia	SCOS type I	182	93.1	5.21	6.86	54.2
10-324	Azoospermia	SCOS type I	183	93.8	4.49	4.97	40.2
15-547	Azoospermia	SCOS type I	187	93	4.59	7.56	50.5
15-368	Azoospermia	SCOS type I	188	88.9	4.77	5.07	51.4

Patient code	Semen Phenotype	Testicular phenotype	Platelets (10^9/L)	MCV (fL)	Erythrocytes (10^{12}/L)	Leucocytes (10^9/L)	Neutrophils (%)
06-183	Azoospermia	SCOS type II	188	88.3	5.3	11.88	64.9
09-263	Azoospermia	SCOS type II	191	88.1	5.61	9.4	60.2
11-417	Azoospermia	SCOS type I	193	86.2	5.15	6.55	58.3
14-277	Azoospermia	SCOS type I	196	88.7	5.13	6.85	52.9
14-313	Azoospermia	SCOS type I	197	92.9	4.79	6.55	60.8
10-393	Azoospermia	SCOS type I	200	85.1	5.74	8.82	54.2

*Patient 14-339 carrying 2 heterozygous pathogenic mutations in *FANCA* gene. Reference value for erythrocytes: 4.3-6.0x10^{12}/L; MCV: 80-100fL; Leucocytes 4.0-10.5x10^9/L; Neutrophils: 50-70%; Platelets: 150-40010^9/L

Table S2. Commercially available TaqMan gene expression assays in testis biopsies.

Gene	Function	Assay ID
FANCA	Target gene	Hs01116668_m1
DAZ	Spermatogonia/early spermatocytes biomarker	Hs00414014_m1
CDY1	Spermatids biomarker	Hs00371514_m1
PRM2	Spermatids/mature spermatozoa biomarker	Hs04187294_g1
BRDT	Pachytene spermatocytes/round and elongating spermatids	Hs00976114_m1
GAPDH	Reference housekeeping gene	Hs0275891_g1

Table S3. Algorithms employed to calculate the index of pathogenicity for variant prioritization.

Algorithm	Prediction	Meaning	Numeric value (u)
SIFT	T	tolerated	-1
	D	damaging	1
	NA	not assigned	0
Polyphen	B	Benign	-1
	P	probably damaging	0.5
	D	damaging	1
	NA	not assigned	0
Mutation Assessor	N	neutral	-1
	L	low	-0.5
	M	medium	0.5
	H	high	1
	NA	not assigned	0
RadialSVM	T	tolerated	-1
	D	damaging	1
	NA	not assigned	0
MutationTaster	P	polymorphism	-1
	D	damaging	1
	NA	not assigned	0
LRT	N	tolerated	-1
	D	damaging	1
	NA	not assigned	0
LR	T	tolerated	-1
	D	damaging	1

Index of pathogenicity (IP) was calculated according to the following formula:

$$\frac{\left(\sum_{r=1}^{n} u_r\right)}{k}$$

Where **k**= number of available prediction tools.

Table S4. Variants detected in *FANCA* gene

Patient code	Mutation	Status	Functional study available	Previously reported
04-170 & brother	NM_000135.2: c.2639C>T (p.Arg880Gln)	Homozygosis	Yes[+]	Reported in 8 FA patients*. Described in monozygotic twins without hematological signs[+]
14-339	NM_000135.2: c.3788_3790delTCT (p.Phe1263del)	Heterozygosis	Yes[£]	Reported in 287 FA patients*.
	NM_000135.2:c.3913C>T (p.Leu1305Phe)	Heterozygosis	Yes[£]	Reported in 5 FA patients*.
14-339's brother (Not affected)	NM_000135.2:c.3913C>T (p.Leu1305Phe)	Heterozygosis	Yes[£]	Reported in 5 FA patients*.

*Fanconi anemia Mutation Database at Rockefeller University (LOVD) https://databases.lovd.nl; [+]Mankad et al. 2006[3]; [£]Castella et al. 2011[4]

Supplementary References:

1 Drmanac R, Sparks AB, Callow MJ, *et al.* Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* 2010; **327**: 78–81.

2 Carnevali P, Baccash J, Halpern AL, *et al.* Computational techniques for human genome resequencing using mated gapped reads. *J Comput Biol* 2012; **19**: 279–92.

3 Mankad A, Taniguchi T, Cox B, *et al.* Natural gene therapy in monozygotic twins with Fanconi anemia. *Blood* 2006; **107**: 3084–90.

4 Castella M, Pujol R, Callen E, *et al.* Origin, functional role, and clinical impact of Fanconi anemia FANCA mutations. *Blood* 2011; **117**: 3759–69.

LICENCE TO PUBLISH

SPRINGER NATURE

Manuscript Number: GIM-D-17-00678R1

Journal Name: Genetics in Medicine (the "Journal")

Proposed Title of the Article: From whole exome analysis in idiopathic azoospermia to the identification of a high risk subgroup for occult Fanconi Anemia (the "Article")

Author(s) [Please list all authors, continuing on a separate sheet if necessary]: Csilla Krausz, Antoni Riera-Escamilla, Chiara Chianese, Daniel Moreno-Mendoza, Elisabet Ars, Osvaldo Rajmil, Roser Pujol, Massimo Bogliolo, Ignacio Blanco, Ines Rodríguez, Isabel Badell, Eduard Ruiz-Castañé and Jordi Surrallés (the "Author(s)")

Miscellaneous [for office use only]:

1. American College of Medical Genetics and Genomics (the 'Licensee') will consider publishing this article, including any supplementary information and graphic elements therein (e.g. illustrations, charts, moving images) (the 'Article'). Headings are for convenience only.

2. Grant of Rights
In consideration of the Licensee evaluating the Article for publication, the Author(s) grant the Licensee the exclusive (except as set out in clauses 3, 4 and 5a) iv) and sub-licensable right, unlimited in time and territory, to copy-edit, reproduce, publish, distribute, transmit, make available and store the Article, including abstracts thereof, in all forms of media of expression now known or developed in the future, including pre- and reprints, translations, photographic reproductions and extensions. Furthermore, to enable additional publishing services, such as promotion of the Article, the Author(s) grant the Licensee the right to use the Article (including the use of any graphic elements on a stand-alone basis) in whole or in part in electronic form, such as for display in databases or data networks (e.g. the Internet), or for print or download to stationary or portable devices. This includes interactive and multimedia use as well as posting the Article in full or in part or its abstract on social media, and the right to alter the Article to the extent necessary for such use. The Licensee may also let third parties share the Article in full or in part or its abstract on social media and may in this context sub-license the Article and its abstract to social media users. Author(s) grant to Licensee the right to re-license Article metadata without restriction (including but not limited to author name, title, abstract, citation, references, keywords and any additional information as determined by Licensee).

3. Self-Archiving
Author(s) are permitted to self-archive a pre-print and an author's accepted manuscript version of their Article.
a) A pre-print is the author's version of the Article before peer-review has taken place ("Pre-Print"). Prior to acceptance for publication, Author(s) retain the right to make a Pre-Print of their Article available on any of the following: their own personal, self-maintained website; a legally compliant, non-commercial pre-print server such as but not limited to arXiv and bioRxiv. Once the Article has been published, the Author (s) should update the acknowledgement and provide a link to the definitive version on the publisher's website: "This is a pre-print of an article published in [insert journal title]. The final authenticated version is available online at: https://doi.org/[insert DOI]"
b) An Author's Accepted Manuscript ("AAM") is the version accepted for publication in a journal following peer review but prior to copyediting and typesetting that can be made available under the following conditions:
i) Author(s) retain the right to make an AAM of their Article available on their own personal, self-maintained website immediately on acceptance,
ii) Author(s) retain the right to make an AAM of their Article available for public release on any of the following 6 months after first publication ("Embargo Period"): their employer's internal website; their institutional and/or funder repositories. AAMs may also be deposited in such repositories immediately on acceptance, provided that they are not made publicly available until after the Embargo Period.
An acknowledgement in the following form should be included, together with a link to the published version on the publisher's website: "This is a post-peer-review, pre-copyedit version of an article published in [insert journal title]. The final authenticated version is available online at: http://dx.doi.org/[insert DOI]"

4. Author's Retained Rights
Author(s) retain the following non-exclusive rights for the published version provided that, when reproducing the Article or extracts from it, the Author(s) acknowledge and reference first publication in the Journal:
a) to reuse graphic elements created by the Author(s) and contained in the Article, in presentations and other works created by them;
b) they and any academic institution where they work at the time may reproduce the Article for the purpose of course teaching (but not for inclusion in course pack material for onward sale by libraries and institutions); and
c) to reproduce, or to allow a third party Licensee to reproduce the Article in whole or in part in any printed volume (book or thesis) written by the Author(s).

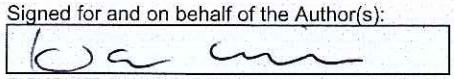
5. Warranties
The Author(s) warrant and represent that:
a) (i) the Author(s) are the sole copyright owners or have been authorised by any additional copyright owner to assign the rights defined in clause 2, (ii) the Article does not infringe any intellectual property rights (including without limitation copyright, database rights or trade mark rights) or other third party rights and no licence from or payments to a third party are required to publish the Article, (iii) the Article has not been previously published or licensed, (iv) if the Article contains materials from other sources (e.g. illustrations, tables, text quotations), Author(s) have obtained written permissions to the extent necessary from the copyright holder(s), to license to the Licensee the same rights as set out in clause 2 but on a non-exclusive basis and without the right to use any graphic elements on a stand-alone basis and have cited any such materials correctly;
b) all of the facts contained in the Article are according to the current body of science true and accurate;
c) nothing in the Article is obscene, defamatory, violates any right of privacy or publicity, infringes any other human, personal or other rights of any person or entity or is otherwise unlawful and that informed consent to publish has been obtained for all research participants;
d) nothing in the Article infringes any duty of confidentiality which any of the Author(s) might owe to anyone else or violates any contract, express or implied, of any of the Author(s). All of the institutions in which work recorded in the Article was created or carried out have authorised and approved such research and publication; and
e) the signatory (the Author or the employer) who has signed this agreement has full right, power and authority to enter into this agreement on behalf of all of the Author(s).

6. Cooperation
a) The Author(s) shall cooperate fully with the Licensee in relation to any legal action that might arise from the publication of the Article and the Author(s) shall give the Licensee access at reasonable times to any relevant accounts, documents and records within the power or control of the Author(s). The Author(s) agree that the distributing entity is intended to have the benefit of and shall have the right to enforce the terms of this agreement.
b) The Author(s) authorise the Licensee to take such steps as it considers necessary at its own expense in the Author(s)' name and on their behalf if the Licensee believes that a third party is infringing or is likely to infringe copyright in the Article including but not limited to initiating legal proceedings.

7. Author List
After signing, changes of authorship or the order of the authors listed will not be accepted unless formally approved in writing by the Licensee.

8. Edits & Corrections
The Author(s) agree(s) that the Licensee may retract the Article or publish a correction or other notice in relation to the Article if the Licensee considers in its reasonable opinion that such actions are appropriate from a legal, editorial or research integrity perspective.

9. Governing Law
This agreement shall be governed by, and shall be construed in accordance with, the laws of the State of New York. The courts of New York, New York shall have the exclusive jurisdiction.

Signed for and on behalf of the Author(s): 

(PLEASE NOTE, ONLY HANDWRITTEN SIGNATURES ARE ACCEPTED)

Print Name: Csilla Krausz

Date: 15/2/2018

Address: Department of Experimental and Clinical Biomedical Sciences " Mario Serio ", Centre of Excellence DeNothe, University of Florence, Viale Pieraccini 6, 50139 Florence, Italy

Corresponding Author Name: Csilla Krausz
Manuscript Number: GIM-D-17-00678R1

Reporting Checklist

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. **Please respond completely to all questions relevant to your manuscript. For sections that are not applicable please fill in NA.** For more information, please read the journal's Guide to Authors.

⬓ Check here to confirm that the following information is available in the Material & Methods section:

- the **exact sample size (*n*)** for each experimental group/condition, given as a number, not a range;
- a **description of the sample collection** allowing the reader to understand whether the samples represent **technical or biological replicates** (including how many animals, litters, culture, etc.);
- a **statement of how many times the experiment shown was replicated in the laboratory**;
- **definitions of statistical methods and measures**: (For small sample sizes (n<5) descriptive statistics are not appropriate, instead plot individual data points)
  - very common tests, such as *t*-test, simple χ^2 tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
  - are tests one-sided or two-sided?
  - are there adjustments for multiple comparisons?
  - **statistical test results**, e.g., ***P* values**;
  - definition of **'center values'** as **median or mean**;
  - definition of **error bars as s.d. or s.e.m. or c.i.**

Please ensure that the answers to the following questions are reported both **in the manuscript itself and in the space below.** We encourage you to include a specific subsection in the methods section each for statistics, reagents and animal models. Below, provide the text as it appears in the manuscript as well as the page number.

Statistics and general methods	**Text AND page number from manuscript**
1. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size? (Give text and page #)	Not applicable
For animal studies, include a statement about sample size estimate even if no statistical methods were used.	Not applicable
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established? (Give text and page #)	Not applicable
3. If a method of randomization was used to determine how samples/animals were allocated to experimental groups and processed, describe it. (Give text and page #)	Not applicable
For animal studies, include a statement about randomization even if no randomization was used.	Not applicable

4. If the investigator was blinded to the group allocation during the experiment and/or when assessing the outcome, state the extent of blinding. (Give text and page #)	Not applicable
For animal studies, include a statement about blinding even if no blinding was done.	Not applicable
5. For every figure, are statistical tests justified as appropriate?	Not applicable
Do the data meet the assumptions of the tests (e.g., normal distribution)?	Not applicable
Is there an estimate of variation within each group of data?	Not applicable
Is the variance similar between the groups that are being statistically compared? (Give text and page #)	Not applicable

Reagents	**Text AND page number from manuscript**
6. Report the source of antibodies (vendor and catalog number)	Not applicable
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination	Not applicable

Animal Models	**Text AND page number from manuscript**
8. Report species, strain, sex and age of animals	Not applicable
9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.	Not applicable

10. We recommend consulting the ARRIVE guidelines (*PLoS Biol.* **8**(6), e1000412,2010) to ensure that other relevant aspects of animal studies are adequately reported.

Human subjects	**Text AND page number from manuscript**
11. Identify the committee(s) approving the study protocol.	Fundació Puigvert
12. Include a statement confirming that informed consent was obtained from all subjects.	Informed consent was obtaine for all subjects
13. For publication of patient photos, include a statement confirming that consent to publish was obtained. For more information, please see http://www.icmje.org/recommendations/browse/roles-and-responsibilities/protection-of-research-participants.html.	Not applicable
14. Report the clinical trial registration number (at ClinicalTrials.gov or equivalent).	Not applicable

15. For phase II and III randomized controlled trials, please refer to the CONSORT statement and submit the CONSORT checklist with your submission.

16. For tumor marker prognostic studies, we recommend that you follow the REMARK reporting guidelines.

Data deposition	**Text AND page number from manuscript**
17. Provide accession codes for deposited data. Data deposition in a public repository is mandatory for: a. Protein, DNA and RNA sequences b. Macromolecular structures c. Crystallographic data for small molecules d. Microarray data	Not applicable

Deposition is strongly recommended for many other datasets for which structured public repositories exist; more details on our data policy are available in the Guide to Authors. We encourage the provision of other source data in supplementary information or in unstructured repositories such as Figshare and Dryad. We encourage publication of Data Descriptors (see Scientific Data) to maximize data reuse.

18. If computer code was used to generate results that are central to the paper's conclusions, include a statement in the Methods section under "**Code availability"** to indicate whether and how the code can be accessed. Include version information as necessary and any restrictions on availability.	Not applicable

COLOR ARTWORK FORM

This form must be completed for all papers. We will be unable to process your paper through to production until we receive instructions concerning color files. Please upload this form with your revised files.

JOURNAL: ***Genetics in Medicine***

ARTICLE TITLE From whole exome analysis in idiopathic azoospermia to the identification of a high risk subgroup for occult Fanconi Anemia

MANUSCRIPT NUMBER GIM-D-17-00678R1

CORRESPONDING AUTHOR NAME Csilla Krausz

Genetics in Medicine has charges for figures printed in color: $500 for the first color figure and $250 for each subsequent figure. Current ACMG members (excluding Student Members) who are first or senior/corresponding authors are exempt,* as are authors who have opted for Open Access.*

Please note that figures can appear online in color in the HTML version of your manuscript, and in black and white in the PDF/print version of the manuscript. Color figures will also be at the discretion of the editorial office.

Please check:

☐ Yes, my manuscript contains material that must be printed in color. I agree to pay the color charges in full and hereby authorize Nature Publishing Group to invoice me for the cost of reproducing color artwork in print.

☒ Yes, my manuscript contains material that should be in color in the online HTML version, but in black and white in the PDF/print version of the manuscript. No charges incurred.

☐ Yes, my manuscript contains material that must be printed in color, but I have chosen Open Access for my article or am an ACMG member, so am exempt.

☐ No, my manuscript does not contain material that must be printed in color.

Which figures should be printed in color? (e.g. Figures 1a, 2, 3b)

_____

College Member's Name: _____

Membership will be verified by the ACMG, and false claims will be liable for the full color charge rate. Signed completion of this form constitutes a full and total acceptance of the terms listed. College membership has no bearing on the review process at *Genetics in Medicine*, and papers from members and nonmembers alike will be given equal consideration.

Color figures will be set close to the citation and in the best possible position.

20% VAT will be added to the total charge amount upon invoicing. This applies to all EU authors who do not provide a valid VAT number upon returning this form. Customers outside of the EU will not be charged VAT but local taxes will be added where applicable.

Signature: _____ VAT no. (if applicable): _____

Print Name: Csilla Krausz Date: 15/02/2018

Updated 8/16/16

nature publishing group npg