SUPP. METHODS

Study Subjects

Genomic DNA samples were from individuals diagnosed with FA and registered in the International Fanconi Anemia Registry (IFAR), following written informed consent. These studies were approved by the Institutional Review Board of the Rockefeller University, New York, USA. The Office of Human Subjects Research at the National Institutes of Health, and Institutional Review Board of the National Human Genome Research Institute (NHGRI) approved the reception of de-identified cell lines and DNA samples from The Rockefeller University and analysis of the underlying molecular variants. DNA extractions from peripheral blood or EBV immortalized cell lines were performed as previously described (Chandrasekharappa, et al., 2013).

Search for mutations

Targeted next generation sequencing, Sanger sequencing, and arrayCGH (array Comparative Genomic Hybridization) were performed as previously described (Chandrasekharappa, et al., 2013; Flynn, et al., 2014). The search for *FANCA* deletions by MLPA (Multiplex Ligation-dependent probe amplification) were performed as previously described (Shukla, et al., 2013). The primers used for Sanger sequencing *SLX4* and *FANCA* mutations were previously reported (Chandrasekharappa, et al., 2013; Flynn, et al., 2014; Kim, et al., 2011).

STRP Genotyping assays

STRP marker genotyping and analysis were performed as previously described (Gunay-Aygun, et al., 2010). The marker labeled XY represents a region of the Amelogenin gene that distinguishes between the X (123 allele) chromosome and the Y (129 allele) chromosome (Sullivan, et al., 1993). The source of the D16S markers is from the "STS markers" track of the UCSC genome browser (http://www.genome.ucsc.edu). The chr16 STRP markers, beginning with "16)", were developed in the lab: the primers were designed around a repeat marker identified in the UCSC browser using the "Microsatellite" track, tested in a number of DNA samples, and found to represent a polymorphic marker. Fragment separation was achieved by capillary electrophoresis on a Genetic Analyzer 3130 using a 36cm capillary array and a POP-7 polymer. The ROX400 (Applied Biosystems) was run as an internal size-standard. Allele size was calculated using the local Southern algorithm available in the GENESCAN software program (Applied Biosystems). Allele calling and binning was done using the GENOTYPER software. (GeneScan and Genotyper are no longer commercially available and have been combined in GeneMapper, Applied Biosystems). All genotyping was done including a CEPH control individual #134702 (Applied Biosystems) for quality control purposes.

SNP arrays and CNV detection

Patient DNA samples were genotyped using the HumanOmniExpress (~750k SNPs) and the HumanOmniExpressExome (~1M SNPs) BeadChip platforms, following the Illumina infinium assay protocol (Gunderson, et al., 2005). The data were scanned by iScan and processed with the GenomeStudio v2011.1 (http://www.Illumina.com)

genotyping module. Samples with a Call Rate < 0.99 and SNPs with a GenTrain score < 0.7 were excluded from analysis. Copy number analysis was performed using CNVPartition v3.2 (Illumina) and Nexus Copy Number v7.5 (Biodiscovery, Inc). The following detection parameters were implemented in CNVPartition to define a CNV: a minimum of 4 contiguous SNPs, a minimum confidence threshold = 0, GC wave adjust = True. The minimum size for detecting regions of homozygosity (ROH) was set at 1Mb. To be considered a CNV, the following criteria had to be met: CN=0 and confidence value>150; CN=1 and confidence value>15; CN=3 and confidence value>25. In Nexus, the SNP-FASST2 Segmentation algorithm was implemented to detect CNVs and ROH, and the default settings were used to define a CNV, with a minimum CNV size set at 1kb and a minimum ROH size set at 1Mb. All CNV and ROH calls were visually evaluated with Nexus v7.5 to remove false positive CNVs and ROH. Mendelian errors were calculated using the "--mendel" command in PLINK v1.07 (http://pngu.mgh.harvard.edu/~purcell/plink/) (Purcell, et al., 2007).

Supp. References

Chandrasekharappa SC, Lach FP, Kimble DC, Kamat A, Teer JK, Donovan FX, Flynn E, Sen SK, Thongthip S, Sanborn E and others. 2013. Massively parallel sequencing, aCGH, and RNA-Seq technologies provide a comprehensive molecular diagnosis of Fanconi anemia. Blood 121(22):e138-48.

Flynn EK, Kamat A, Lach FP, Donovan FX, Kimble DC, Narisu N, Sanborn E, Boulad F, Davies SM, Gillio AP, 3rd and others. 2014. Comprehensive Analysis of Pathogenic Deletion Variants in Fanconi Anemia Genes. Human mutation.

Gunay-Aygun M, Zivony-Elboum Y, Gumruk F, Geiger D, Cetin M, Khayat M, Kleta R, Kfir N, Anikster Y, Chezar J and others. 2010. Gray platelet syndrome: natural history of a large patient cohort and locus assignment to chromosome 3p. Blood 116(23):4990-5001.

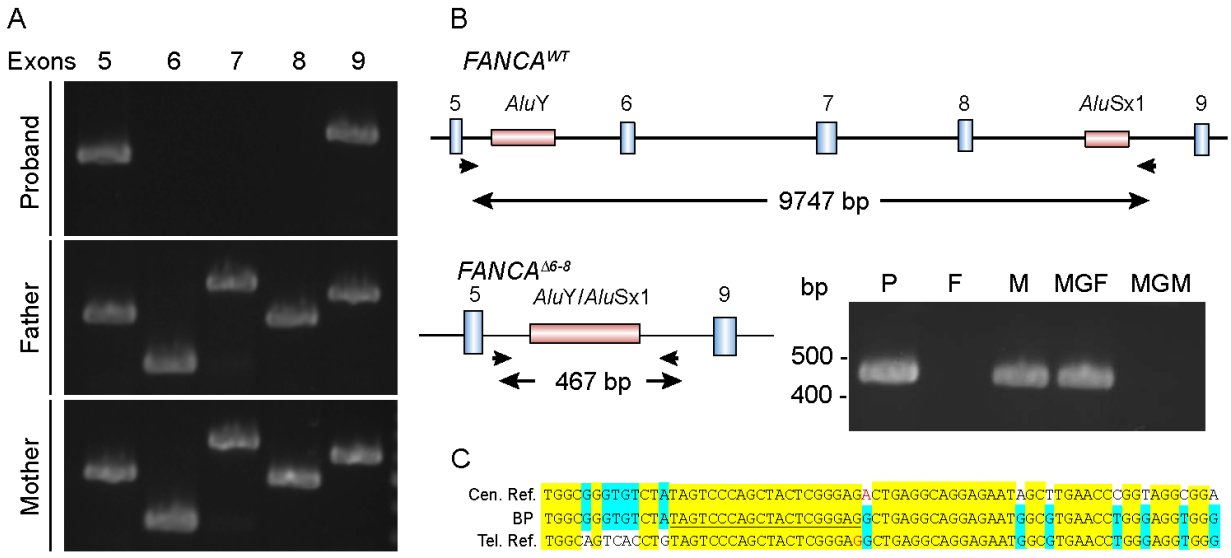
Gunderson KL, Steemers FJ, Lee G, Mendoza LG, Chee MS. 2005. A genome-wide scalable SNP genotyping assay using microarray technology. Nat Genet 37(5):549-54.

Kim Y, Lach FP, Desetty R, Hanenberg H, Auerbach AD, Smogorzewska A. 2011. Mutations of the SLX4 gene in Fanconi anemia. Nat Genet 43(2):142-6.

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

Shukla P, Rao A, Ghosh K, Vundinti BR. 2013. Identification of a novel large intragenic deletion in a family with Fanconi anemia: first molecular report from India and review of literature. Gene 518(2):470-5.

Sullivan KM, Mannucci A, Kimpton CP, Gill P. 1993. A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X-Y homologous gene amelogenin. Biotechniques 15(4):636-8, 640-1.



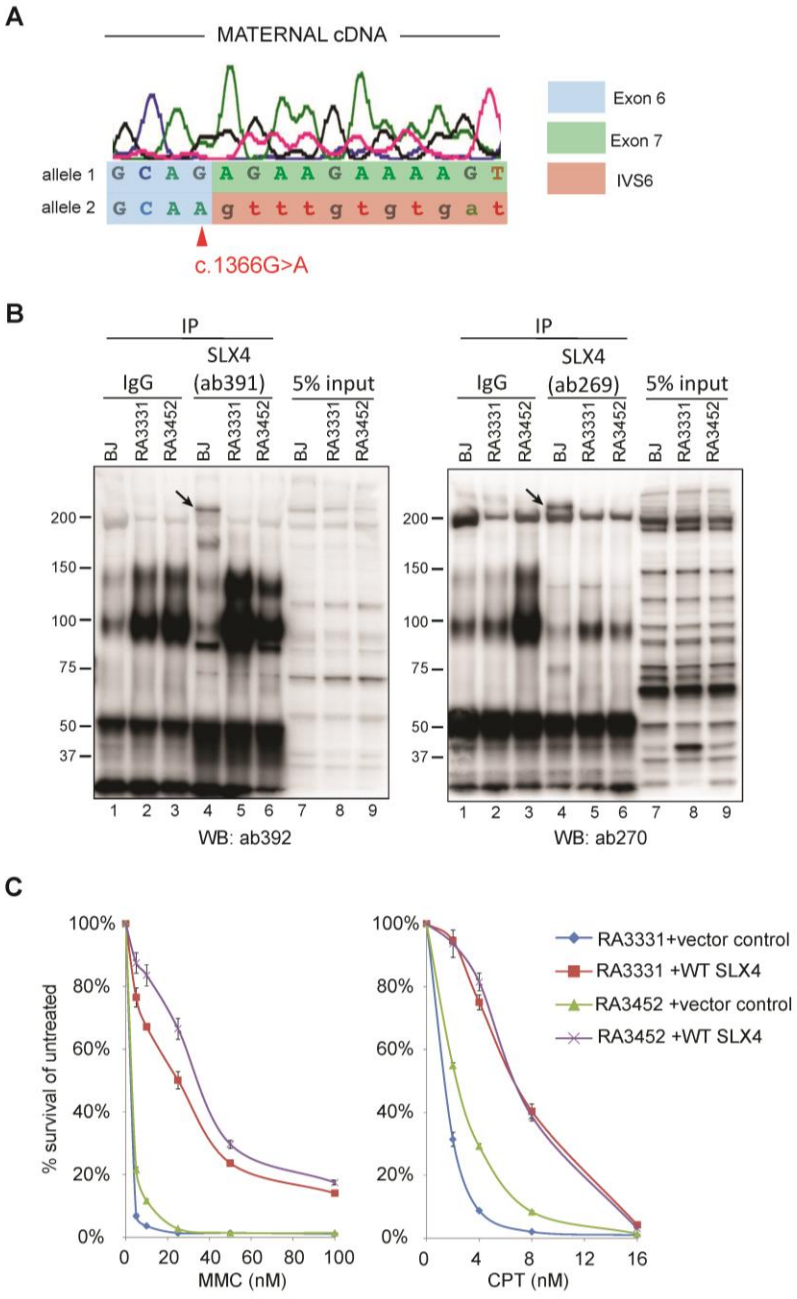
Supp. Figure S1. Deletion of *FANCA* exons 6-8 in patient FA3

A. *PCR products, FANCA exons 5-9, from FA3 proband and parental genomic DNA*. No PCR products are present for exons 6-8 in the proband, indicating a homozygous deletion of these exons. The primers used in the PCR of each exon, and the product size, are as follows: exon 5: CCCTTTTTCATCCACTCTCTG; TGGTTTTAGATTTTGTCTCATGTGT (599bp); exon 6: GCCCAAAGGCCAGGGAGTTT; AGCCAGAAATCAAACCCGTCTGA (416bp); exon 7: TCCACATCAGTCAAGTAAAATGC; GGGAATTGGGAAAATCCAC (752bp); exon 8: TGGAGGTGACGGAGACTCTAA; AAAATCTGACTATAAACTTGTTCCCTTT (585bp); exon 9: CAAGGGCATTCCAGTACCAG; TCTTAGGGTAATTCCAGAAAGTTCA (697bp). The primers had M13 tags and the product size includes the length of these tags.

B. *Mapping the deletion breakpoint in FA3*. The *FANCA* gene region spanning exons 5-9, the location of the primers (small arrows) used for breakpoint PCR, and the product

sizes of the wild-type (WT) and mutant (Δ6-8) alleles are shown. The deletion breakpoints occur in *Alu*Y and *Alu*Sx SINE elements at the telomeric (tel) and centromeric (cen) ends of IVS5 and IVS8, respectively. PCR primers were designed to amplify the breakpoint in the genomic DNA, producing a product size of 467bp: AACATACACACAGACCAACGGAACAGG, and CGTTTAGCCTTGGGTATATTAGTATGGAG. The gel image shows the deletion is present in the proband (P) and the mother (M), but absent from the father (F). The deletion is present in the maternal grandfather (MGF), but not present in the maternal grandmother (MGM).

C. *Sequence alignment of the breakpoint junction sequence*. The genomic breakpoint junction sequence of FA3 (BP) and its alignment with the centromeric and telomeric reference sequences (Cen. ref and Tel. ref) are shown. Identical base pairs in all three sequences are highlighted in yellow, identical base pairs in two of the three sequences are highlighted in blue, and the overlapping sequence where the breakpoint occurs is underlined. The first and last nucleotides of the deletion are marked in red (centromeric nucleotide: chr16:g.89867339 and telomeric nucleotide: chr16:g.89876628 (hg19)).



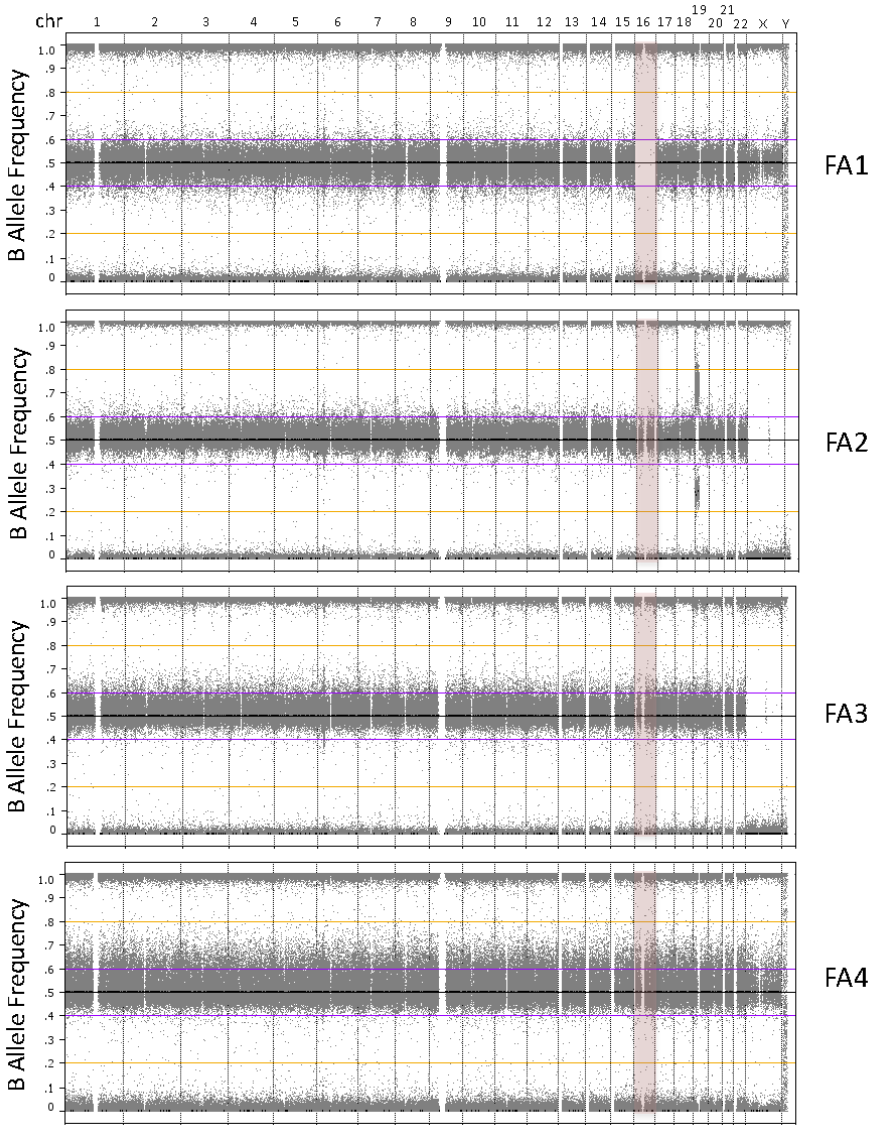
Supp. Figure S2. Pathogencity of *SLX4/FANCP* variant, c.1366G>A

A. Sequencing of the cDNA derived from the maternal lymphoblasts. Junction of Exon 6 (blue) and Exon 7 (green) is shown. Novel c.1366G>A variant results in aberrant splicing and retention of IVS6 (red) in the cDNA. The allele carrying the variant is

unstable due to a stop codon 102 nucleotides later thus the sequencing signal of that allele is very low. Theoretical translation of that allele would lead to expression of p.E456Sfs*34 SLX4, which would remove the domain necessary for binding to XPF. So this allele, even if expressed, would not be functional. We were unable to amplify the full length transcript from the proband's fibroblasts, which is consistent with nonsense mediated decay of the mRNA carrying the variant.

B. Patient's fibroblasts (RA3452) do not express SLX4. Western blot analysis with antibody 392 and 270 against SLX4 after immunoprecipitation from whole cell extracts with antibody 391 (left) and antibody 269 (right). Antibody 391 and 392 (developed in-house) recognizes N terminus and antibody 269 and 270 (Bethyl Laboratories, Inc) recognize the C terminus of SLX4. SLX4 protein is indicated by a black arrow and is present in BJ (WT control cells) and absent from SLX4-null fibroblasts RA3331. 392 and 270 antibodies are unable to specifically recognize SLX4 in the input. All the input bands are nonspecific.

C. Proband's fibroblasts (RA3452) are sensitive to MMC, and CPT and the sensitivity is rescued with expression of WT SLX4. SLX4 null cells (RA3331) are used as control. Sensitivity is rescued with overexpression of WT SLX4. All fibroblast cell lines are transformed and immortalized by expression of E6E7 and hTERT.



Supp. Figure S3. Genome-wide SNP genotyping data for patients displaying UPD. The figure shows the B Allele Frequency (BAF) plots for each UPD patient: FA1, FA2, FA3, and FA4. The BAF conveys the zygosity of each SNP. At each position, a SNP's genotype can either be homozygous (a BAF value of 1 or 0), or heterozygous (a BAF value of 0.5). The genome wide plots show that the only relatively large stretches of homozygosity occur in chromosome 16. The Log R Ratio plots indicate disomy throughout the genome (data not shown).

Supp. Table S1. Genomic STRP Markers and STRP Analysis of UPD Families

		FA1			FA2			FA3			FA4		
Chr	**Marker ID**	**Proband**	**Father**	**Mother**	**Proband**	**Father**	**Mother**	**Proband**	**Father**	**Mother**	**Proband**	**Father**	**Mother**
2	D2s2739	341/355	345/341	335/355	335/329	339/335	335/329	317/345	335/317	317/345	331/328	317/331	345/328
3	D3s3571	254/254	269/254	254/254	265/259	265/265	259/259	264/280	264/264	264/280	262/256	259/262	278/256
4	D4s402	315/343	333/315	345/343	309/309	309/309	341/309	309/309	309/309	341/309	311/323	321/311	311/323
6	D6s1613	120/118	126/120	118/118	126/114	136/126	118/114	118/124	124/118	130/124	126/128	118/126	134/128
8	D8s537	188/180	192/188	189/180	194/192	194/194	174/192	192/195	195/192	188/195	192/195	192/192	198/195
9	D9s1799	186/185	186/186	181/185	180/191	175/180	197/191	187/197	161/187	197/197	195/192	181/195	177/192
11	D11s922	117/117	147/117	119/117	127/143	113/127	119/143	147/113	113/145	121/113	135/119	119/135	131/119
12	DS12s78	210/208	200/210	196/208	206/194	194/206	196/194	220/208	196/220	212/208	208/214	214/208	214/214
15	D15s211	272/264	266/272	254/264	252/228	266/252	250/228	277/252	250/277	258/252	268/262	228/268	262/262
18	D18s488	261/261	261/261	277/261	-	-	-	-	-	-	267/261	-	265/261
22	D22s283	149/167	161/149	159/167	167/157	165/167	171/157	163/153	159/163	167/153	159/153	155/159	151/153
XY	Amelogenin*	123/123	129/123	123/123	129/123	123/129	123/123	129/123	123/129	123/123	123/123	129/123	123/123

*the primers generate 123bp and 129bp products for the Amelogenin locus on chrX and chrY.

red font: indicates the allele that was inherited by the proband, and conveys that both parents show a contribution of alleles.

Supp. Table S2. Chromosome 16 STRP markers used in each UPD family

Marker ID	Chr	Position*	FA1	FA2	FA3	FA4
D16s3401	16	231931				x
D16s3124	16	2447563				x
D16s2617	16	3252451		x	x	x
D16S423	16	6043141	x	x	x	x
D16S404	16	9718112				x
D16S3075	16	12209198		x	x	x
D16S3046	16	20886398				x
D16S3068	16	25560601		x	x	
D16s690	16	27958634	x	x	x	x
D16s3022	16	28320036		x	x	
D16s753	16	31273449		x	x	
D16s746B	16	34858004	x	x	x	
D16s3044	16	47437536		x	x	
D16s3080B	16	49683350		x	x	
D16S3136	16	50706233	x	x		x
D16s419	16	52953362		x	x	
D16s3140C	16	56311663		x	x	
D16S503	16	63598788	x	x	x	
D16S397	16	66738209		x	x	
D16S3018	16	74172611		x	x	
D16S516	16	79124016		x	x	
D16S3091	16	82980450		x	x	
D16S520	16	86516112		x		x
16)87919297	16	89391796		x	x	
D16s3026	16	89492878	x			
D16s3121	16	89498593	x			
16)88222563	16	89695062		x	x	
16)88470941	16	89943440		x	x	

*coordinates are in accord with UCSC genome build hg19
red font: results for these markers were informative and are presented in Figure 1
blue line = position of FANCP/SLX4: hg19_g.chr16.3,631,184-3,661,585
orange line = position of FANCA: hg19_g.chr16.89,803,959-89,883,065

Supp. Table S3. Regions of Uniparental Isodisomy in Four FA Patients

Patient	Chr	Start*	End	Length	Chromosome Terminus	Disease-Causing FA Gene
FA1	16	1	90163275	90163275	p, q	*FANCA*
FA2	16	1	13217794	13217794	p	
		77686375	90163275	12476900	q	*FANCA*
FA3	16	1	13845708	13845708	p	
		84645486	90163275	5517789	q	*FANCA*
FA4	16	1	8711547	8711547	q	*SLX4*

*coordinates are in accord with UCSC genome build hg19