Identification and Initial Characterization of The Fanconi Anemia Complementation Group D Gene

By
Cynthia D. Timmers

A DISSERTATION

Presented to the Department of Molecular and Medical Genetics
and the Oregon Health Sciences University School of Medicine
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
April 1999

School of Medicine
Oregon Health Sciences University

CERTIFICATE OF APPROVAL

This is to certify that the Ph.D. thesis of

Cynthia D. Timmers

has been approved

[redacted]

Professor in charge of thesis

[redacted]

Member

[redacted]

Member

[redacted]

Member

[redacted]

Member

Associate Dean for Graduate Studies

Table of Contents

List of Illustrations

Tables

Figures

Acknowledgments

It is very difficult to write this section because so many people have played an influential role in helping me attain this goal. So I would like to begin by noting that this work represents a huge collaborative effort on the behalf of many people not only on a physical level but also on the intellectual and emotional levels.

First and foremost, I extend my deepest gratitude to my advisor, Markus Grompe. You not only gave me the opportunity to prove myself scientifically, but encouraged and supported me throughout the process. You were always there playing the optimist when things looked bad and the pessimist when I got overly excited. I only hope that I can live up to your expectations for me. I must also thank you for the endless supply of free beers at "the pub" and commend you on your ability to communicate with the lab on a personal level. I knew you were the perfect advisor for me, when we showed up in San Francisco for Halloween wearing the same set of plastic boobs!

Next, I would like to thank all the members of the Grompe lab, both present and past, who have had to endure my somewhat militant behavior. You have been a great group to be around both at work and play. I would especially like to thank Mike Whitney for taking me under his wing when I first started and teaching me all the quick and dirty ways to do molecular biology. And of course who could forget all those "boating" days! I would also like to thank Yassmine Akkari (The Little One). Looking back at it now, I cannot believe the amount of time you spent helping me write this thesis and prepare for

biggest supporter of all has been my best friend and boyfriend, Brian. You are probably one of the main reasons I am writing this today, not only because you made me come to work every blasted weekend for the last three years but also because you stood by me through the thick and the thin. You taught me what the word relationship really means and no matter what the future has to hold for the two of us, there will always be a special place for you in my heart.

Finally, I would like to dedicate this thesis to my grandma. You have spent countless hours throughout my life listening to my problems and giving me advise. You taught me that nothing was impossible if you put your mind into it and that I should never settle for second best. I am sorry we haven't talked as much since I moved to Portland, but I want you to know that I think of you often and remember all the advise you have given me. So Grams, I may not have found a cure for cancer but I have taken the first step in gaining the knowledge to maybe some day helping figure it out!

my seminar. You have got to be one of the most generous and patient people on this earth considering you put up with me! I am so thankful you took the time to look beyond my tough outer-shell. You have truly become one of my closest friends and although I will miss you when I leave, I am sure we will always stay in touch.

I would like to thank my thesis committee, Drs. Matt Thayer, Susan Olson, Karen Rodland and Elaine Lewis for their excellent advise and encouragement. I would especially like to thank Matt for his open door whenever I had a "quick" question or exciting news to share. And of course, thank you for all the six packs of beer I won betting on the Packers against the 49ners. Although I had to buy the beer the last time, we all know who really won that game!

Last but not least, I would like to thank all my friends and family. Although many of you have no idea of why I have been in school for so long or what exact field I am getting my degree in, you have been supportive and encouraging. Of course my

Abstract

Fanconi anemia (FA) is an autosomal recessive disorder characterized by a variety of congenital abnormalities, bone marrow failure and an increased risk for malignancy. FA cells exhibit chromosomal instability and are hypersensitive to DNA cross-linking agents . FA is a clinically heterogeneous disorder and can be functionally divided into at least eight different complementation groups (*FANCA-H*). Three of the eight genes (*FANCA, FANCC and FANCG*) have been identified using an expression cloning technique. A fourth gene, *FANCD*, has been mapped to chromosome 3p22-26.

The first section of this thesis describes the sub-localization of *FANCD* to a 200 kb region within the previously-defined location. We used microcell-mediated chromosome transfer to fuse a neo-marked human chromosome 3 into a $FANCD^-$ patient cell line, PD20. It had been previously demonstrated that this technique could be used to correct the cellular hypersensitivity to DNA cross-linking agents as determined cytogenetically by chromosome breakage studies. To sub-localize the gene, we chose to analyze hybrids which remained sensitive to these agents even though they had received a donor chromosome 3, theorizing that the donor chromosome had incurred a deletion of the *FANCD* gene and was, therefore, no longer able to correct the cells. By defining the location of the deletions in five individual hybrids, we were able to narrow the region where the gene must lie to approximately 200 kb.

To facilitate the sub-localization of *FANCD*, a BAC contig was constructed covering approximately 1-2 Mb on chromosome 3p25.3. The contig enabled us to identify three candidate genes that mapped to the above region. Two of the genes were eliminated as candidates based on their expression pattern, and mutation analysis was performed on the third, SGC34603. Two non-polymorphic sequence alterations were identified in the PD20 cell line. Additional evidence for the gene was obtained through the analysis of a recent bone marrow sample from the PD20 patient which appeared to contain a sub-population of reverted cells. By analyzing cDNA from this sample we showed that a mitotic intragenic recombination event had occurred in the lymphoblast of the patient. Selection for this type of an event in SGC34603, would only occur if this gene was indeed FANCD. Based on these results, we now believe that SGC34603 is the *FANCD* gene.

The *FANCD* gene has an open reading frame of 4,413 bp comprised of 43 exons and encodes a 1,472 amino acid protein. The protein does not share any functional homology with any other known proteins, but interestingly, appears to have a possible homolog in *Arapodopsis thaliana.* Unlike the other FA genes, *FANCD* is toxic when over-expressed in cells. It would, therefore, have been impossible to identify it using the expression cloning method.

Additionally, a murine homolog for *FANCD* was cloned in preparation for the construction of a mouse knock-out.

Chapter One:
Introduction

I. Clinical Characteristics of Fanconi Anemia

Fanconi anemia (FA) was first described in 1967 by a Swiss pediatrician, Guido Fanconi (Fanconi, 1967). It is a rare, autosomal recessive disorder clinically characterized by developmental defects, bone marrow failure, and cancer susceptibility. The estimated incidence is 1:100,000 in the general population. Although FA is observed worldwide, there is a higher incidence in certain ethnic groups including the Ashkenazi Jews, South Africaners and in certain Turkish and Saudi Arabian populations (Rosendorff, et al., 1987, Verlander, et al., 1995). Somatic cell fusion studies have revealed the existence of at least eight different complementation groups, termed Fanconi anemia complementation group A-H (FANCA-H) (Joenje, et al., 1997).

The severity and prevalence of the developmental defects in FA is highly variable, with one or more abnormalities being present in approximately 50 percent of patients (Evans, et al., 1994). Patients may present with multiple congenital defects including growth retardation, thumb and radial defects, kidney and renal malformations, skin hyperpigmentation, microcephaly, congenital heart defects, and mental retardation (Glanz and Fraser, 1982, Chaganti and Houldsworth, 1991, Alter, 1993). Male patients may also have underdeveloped gonads and abnormal spermatogenesis. The wide range of organs affected in FA suggests that the genes involved are required in a general developmental process during normal human embryogenesis.

Typically, during the first decade of life, FA patients will develop macrocytosis and pancytopenia (Alter, et al., 1991, Gordon-Smith and Rutherford, 1991, Butturini, et

al., 1994, Alter, 1995). The average age of onset is 7 years with most FA cases being diagnosed between the ages of 3 and 10 (Butturini, et al., 1994, Joenje, et al., 1995b). There are, however, examples of people who have been diagnosed well outside these ranges (Liu, et al., 1991). The development of pancytopenia represents an inherent bone marrow defect in FA, which results in the patient's inability to produce sufficient amounts of all peripheral blood lineages. Deficiencies in platelets or red blood cells usually precede white blood cell abnormalities. The hematological abnormalities usually present clinically as susceptibility to infection, anemia, or clotting problems. The mean age of survival in FA is 16 years with death usually resulting from complication of bone marrow failure.

Patients with FA are also prone to the development of malignancies. Approximately 52 percent of patients surviving the anemia, develop acute myeloblastic leukemia by age 40 (Butturini, et al., 1994). Cancers in several other organs including skin, gastrointestinal and gynecological systems, have also been described (Swift, et al., 1971, Nara N., 1980, Somers, et al., 1995, Alter, 1996).

Diagnosing FA on the basis of clinical manifestations alone is difficult since the variable phenotype overlaps considerably with several other genetic and non-genetic diseases (Auerbach, 1993). A cellular assay has therefore been developed that exploits the sensitivity of FA cells to bifunctional alkylating agents. FA cells exhibit elevated levels of spontaneous chromosomal breakage and radial formations that are amplified by the addition of cross-linking agents, such as diepoxybutane (DEB) or mitomycin C (MMC). Although similar spontaneous chromosomal changes are seen in Bloom

syndrome and ataxia telangiectasia, the damage is not DEB-induced in these diseases (Hojo, et al., 1995, Meyn, 1995). Therefore, the DEB test is highly sensitive and specific for FA.

The diagnosis of FA may also be made more difficult by the presence of somatic mosaicism. A recent study reported that approximately 25% of FA patients have evidence of spontaneously occurring mosaicism as manifested by the presence of two populations of lymphocytes, one which is hypersensitive to cross-linking agents, and one that behaves normally (Lo Ten Foe, et al., 1997). Although the molecular basis for the phenotypic reversion may vary, it has been demonstrated that intragenic mitotic recombination can occur. In fact, in the same study listed above, researchers found that the mosaicism in one patient was the result of an intragenic crossover between the maternally and paternally inherited mutations. In two others, the mechanism of reversion appeared to have been gene conversion resulting in segregants which had lost one pathogenic mutation. Mosaicism due to mitotic recombination has also been described for other diseases including Bloom syndrome and adenosine deaminase deficiency (Ellis, et al., 1995, Hirschhorn, et al., 1996).

Currently, the best treatment for FA is bone marrow (BM) transplantation, especially if a related, HLA-matched donor is available. In one study (Kohli-Kumar, et al., 1994), eighteen FA patients with allogeneic donors underwent BM transplantation. Seventeen patients sustained engraftment and remained transfusion-independent. Unfortunately, several drawbacks to this procedure exist. First, because of the use of DNA-damaging agents such as cyclophoshamide, as part of the BM transplantation

preparative regimen, the risk of developing leukemia or other cancers after transplantation is high. Second, it is necessary to use a lower dose of preparative agents, because FA patients are more sensitive to them (Soci'e, et al., 1993, Flowers, et al., 1996). This can cause a problem especially when myelodysplastic syndrome is already present, since the lower doses are usually insufficient to kill all the pre-leukemic cells. Finally, FA patients often develop severe toxicity from graft-versus-host disease (GVHD), perhaps as a result of an increased cellular sensitivity to endogenous cytokines released during GVHD.

For patients lacking a fully matched donor, treatment is usually limited to supportive care for their BM failure. In approximately 50 percent of FA patients, treatment with androgens provides improvement in peripheral blood counts (Young and Alter, 1994), but even those patients who initially respond well, become refractory to the treatment after time, and eventually require erythrocyte and platelet transfusions to maintain adequate peripheral blood counts. Treatment with hematopoietic growth factors (GM-CSF and G-CSF) may give an improvement in white blood count, but do not generally improve erythrocyte or platelet counts (Guinan, et al., 1994, Rackoff, et al., 1996). Bone marrow transplantation from alternative donors has been attempted for patients who lack HLA-identical siblings, but the results have been poor with very few long-term survivors (Davies, et al., 1996). Transplantation with hematopoietic stem cells from bone marrow or umbilical cord blood is also an option for patients without a matched donor (Broxmeyer, et al., 1989, Broxmeyer, et al., 1990, Wagner, 1993, Kurtzberg, 1996, Kurtzberg, et al., 1996).

Finally, with the cloning of several FA genes, gene therapy has become an option to cure the aplastic anemia in patients who lack a HLA-matching donor (Liu, 1998). In vitro correction of the FA cellular phenotype has been shown in both lymphoblast and $CD34^+$-selected hematopoietic progenitor cells using either a recombinant retrovirus or adeno-associated viral vector (Walsh, et al., 1994a, Walsh, et al., 1994b). It is hypothesized that transduction of FA progenitor and stem cells might lead to a sustained production of genetically corrected hematopoietic cells of all lineages. The corrected hematopoietic cells are expected to have a growth advantage and should expand, presumably leading to a repopulation of the bone marrow (Liu, et al., 1994). A phase I clinical trial has been performed using the FANCC gene in a retroviral vector (Fanconi Anemia Symposium, Baltimore, MD, 1997). In this study, hematopoietic progenitor cells from patients were harvested, transduced ex-vivo with a retroviral vector containing the wild-type FANCC gene and then reinfused into the patient. The treatment was repeated every three months over a year. Initial results looked promising, however, after a year, the patient did not maintain the corrected cells. One possibility is that the efficiency of the transduction using retroviral-based vectors is to low and was unable to transduce stem cells. Trials with other vectors may prove to be more successful.

II. Cellular Phenotypes of Fanconi Anemia

The hallmark phenotype of Fanconi Anemia is the cellular hypersensitivity to interstrand DNA cross-linking agents (Sasaki and Tonomura, 1973, Auerbach and Wolman, 1976). The sensitivity is specific to agents such as DEB and MMC that predominately produce interstrand cross-links. In contrast, FA cells do not show increased sensitivity to X-rays or monofunctional DNA alkylating agents such as methyl methane sulfonate (MMS) or ethyl methane sulfonate (EMS) (Duckworth-Rysiecki and Taylor, 1985, Kupfer and D'Andrea, 1996, Johnstone, et al., 1997). FA cells exhibit chromosomal instability, which is observed in the metaphase preparation as gaps, breaks, radial exchange figures and endoreduplication (Cohen, et al., 1979). The rate of these chromosomal defects is dramatically increased when FA cells are exposed to DNA cross-linking agents. Several other aberrant processes in FA cells have also been described, including a sensitivity to oxygen, a G2 delay in response to cross-linking agents, and an abnormal production of certain growth factors (Weksberg, et al., 1979, Saito and Moses, 1992, Berger, et al., 1993, Seyschab, et al., 1993). Because the underlining defect in FA is unknown, it remains to be established which of these features are primary to the disease and which are merely secondary manifestations.

Historically, Fanconi anemia has been categorized as a DNA repair defect because of the hypersensitivity of FA cells to DNA cross-linking agents. FA was considered one of the sister diseases to xeroderma pigmentosum (XP), characterized by sensitivity to UV light, and ataxia telangiectasia (AT), similarly sensitive to ionizing

radiation and radiomimetic drugs (Buchwald and Moustacchi, 1998). In light of this classification, a considerable effort has been aimed at identifying a fundamental DNA repair defect through the analysis of FA cells. An initial study, using *FANCA* $^-$ fibroblasts and lymphoblasts, showed that DNA synthesis does not recover following treatment with 8-methyl-oxy-psoralen (8-MOP) and ultraviolet A (UVA) (Moustacchi, et al., 1987). However, this does not appear to be the same in *FANCB* $^-$ or *FANCD* $^-$ cell lines (Moustacchi, et al., 1987, Digweed, et al., 1988). A subsequent study reported that removal of damaged DNA following 8-MOP and UVA as measured by unscheduled DNA synthesis (UDS), is deficient in *FANCA* $^-$ and *FANCB* $^-$ cells as compared to controls (Lambert, et al., 1992). It was then reported that *FANCA* $^-$ cells lack a damage recognition protein which binds to DNA containing interstrand cross-links (Hang, et al., 1993). It is currently believed that the damage recognition protein is part of a complex that binds DNA and with the use of an endonuclease, incises the cross-linked DNA. Finally, it has been demonstrated that extracts of *FANCA* $^-$ cells can correct the UDS defect in *FANCD* $^-$ cells and vice versa (Lambert, et al., 1997). In contradiction to this work, experiments done with *FANCC* $^-$ lymphoblasts versus controls showed that the difference is in the initial formation of cross-links, and that if one compares removal of cross-links, no differences can be detected (Youssoufian, 1996). These results, however, were obtained using MMC and were not observed when repeated with psoralen plus UVA (Y. Akkari, personal communication). A possible explanation for this difference could be that MMC only effectively induces cross-links at low levels of oxygen (5%) and is inactivated by normal levels (20%) (Pritsos and Sartorelli, 1986). Since most

Table 1. Cellular Abnormalities of Fanconi Anemia

	References
Sensitivity to Cross-linking Agents	(Schroeder, et al., 1964) (German and Puglatta, 1966)
Prolongation of G2 Phase of Cell Cycle	(Kaiser, et al., 1982) (Kubbies, et al., 1985)
Sensitivity to Oxygen	
-Grow poorly at ambient 0_2	(Schindler and Hoehn, 1988)
-Overproduction of O2 radicals	(Korkina, et al., 1992)
-Deficient 02 radical defense	(Gille, et al., 1987)
-Deficient in superoxide dismutase	(Joenje, et al., 1979) (Mavelli, et al., 1982)
Sensitivity to Ionizing Radiation (G2 specific)	(Bigelow, et al., 1979)
Overproduction of TNF-α	(Rosselli, et al., 1992)
Direct Defects in DNA Repair	
-Accumulation of DNA Adducts	(Takeuchi and Morimoto, 1993)
-Defective in repair of DNA cross-links	(Fujiwara and Tatsumi, 1977)
Genomic Instability	
-Spontaneous Chromosome Breakage	(Auerbach, et al., 1989)
-Hypermutable (by deletion mechanism)	(Papadopoulo, et al., 1990a)
Increased Apoptosis	(Willingale-Theune, et al., 1989) (Kupfer and D'Andrea, 1996) (Wang, et al., 1998)
Defective p53 induction	(Rosselli, et al., 1995) (Kupfer and D'Andrea, 1996)
Intrinsic stem cell defect	
-Decreased colony growth in vitro	(Daneshbod-Skibba, et al., 1980) (Alter, et al., 1991)
-Decreased gonadal stem cell survival	(Whitney, et al., 1996)

(Adapted with permission from (D'Andrea and Grompe, 1997))

experiments studying the effects of MMC on FANC cells are done at 20% oxygen, the bulk of the DNA damage is probably due to the high levels of free radicals produced by the cycles of oxidation and reduction of MMC and not from cross-links (Clarke, et al.,

1997). It could also be that the primary defect in FA is not the DNA repair process itself, but some other process that modulates the amount of damage and/or the fate of the damage. In this case, each FANC protein could function in separate aspects of the disease, which would explain the differences in the results.

Since it is generally believed that mutagenic events are the result of inaccurate repair steps, several mutagenesis studies have been done in hopes of finding a difference between FA and control cells so that one may deduce the aberrant step. Through experiments using normal human lymphoblasts, it has been shown that the majority of spontaneous and psoralen-photoinduced mutations in the hypoxanthine phosphoribosyl-transferase (*HPRT*) locus, are base pair substitutions. In contrast, in FA lymphoblasts the majority of mutants are the result of deletions in the *HPRT* coding sequence (Papadopoulo, et al., 1990a, Papadopoulo, et al., 1990b, Guillouf, et al., 1993, Papadopoulo, et al., 1993). In support of this data, a study using erythrocytes from FA patients showed an increased frequency of glycophorin A (*GPA*) mutants, a locus most commonly inactivated by deletions (Sala-Trepat, et al., 1993). It has, therefore, been hypothesized that the observed increase in deletions in response to the DNA cross-linking agents may be caused by the inability of FA cells to repair double strand breaks. The best evidence to support this theory comes from a study using a DNA end-joining assay in cultured cells (Escarceller, et al., 1998). The results of this experiment showed that although the overall ligation efficiency of double strand breaks (DSB) was normal in FA lymphoblasts, error-free processing of blunt-ended DSB was markedly decreased in *FANCC* $^{-}$ cells, resulting in a higher deletion frequency and larger deletion size. This

phenotype could be completely complemented with the introduction of the wild-type *FANCC* gene. In addition to this study, it has been reported that FA cells have a substantially elevated level of homologous recombination as compared to controls (Thyagarajan and Campbell, 1997). Taken together, these results provide evidence that FA genes may play a role in the fidelity of end-joining DNA. This model would explain the genomic instability and cancer predisposition characteristic of the disease. Unfortunately, little is known about the mechanistic relationship between fidelity of DSB end-joining, deletion events, and other cellular processes such as cell cycle control or processing of cross-linked DNA. It is possible that the FA genes could be involved in any of these processes and still yield the same results.

Primary lymphoblast and fibroblast cultures derived from FA patients have been shown to grow poorly under standard oxygen conditions (20%) (Weksberg, et al., 1979). As mentioned previously, many DNA cross-linking agents are known to produce oxygen reactive species. Therefore, one possible defect that would account for the sensitivity of FA cells to these agents could be the inability of the cells to properly respond to such reactive species. Evidence to support this idea includes the fact that cultured fibroblasts from FA patients exhibit strong oxygen hypersensitivity as monitored by serial passaging and cell cloning experiments (Schindler and Hoehn, 1988). It has also been shown that FA patients have a higher intracellular level of reactive oxygen intermediates as well as an increase in an oxidative DNA lesion, 8-hydroxydeoxyguanine (8-OhdG) (Takeuchi and Morimoto, 1993, Ruppitsch, et al., 1997). On the other hand, introduction of the *FANCC* cDNA into FA lymphoblasts does not alter their oxygen sensitivity (Joenje, et

al., 1995c). It has also been shown that only primary and not SV40 transformed FA fibroblasts are sensitive to oxygen, even though the transformed cells retain their hypersensitivity to DNA cross-linkers (Saito, et al., 1993). These results would argue that the oxygen sensitivity is a secondary feature of the FA defect.

Patients with Fanconi anemia typically develop pancytopenia during the first decade of life. Because of the development of this bone marrow defect, a plethora of studies have looked at the levels of growth factors involved in the proliferation of BM cells. Several observations suggest that FA cells produce less interleukin-6 (IL6) than controls and have increased levels of tumor necrosis factor-α (TNFα) (Bagnara, et al., 1992, Rosselli, et al., 1992, Rosselli, et al., 1994). It has also been shown that adding IL6 or anti-TNFα antibodies to cells can reduce the chromosomal instability induced by MMC (Rosselli and Moustacchi, 1990, Rosselli, et al., 1992). Studies looking at the levels of these two growth factors in patient samples are not as clear. Many of the studies pooled patients from different complementation groups, which would explain the differences in results. These studies do, however, provide some clues to the basic alterations in the biology of FA cells and can now be more accurately assessed using mouse models.

The G2/M transition is genetically regulated in response to DNA damage (Hartwell and Weinert, 1989, Weinert, et al., 1994). It serves as a checkpoint to delay cell cycle progression in order to allow the repair of damaged DNA. Mutations in the proteins involved in this regulation could result in genomic instability because the cells would enter mitosis prematurely. Several studies have aimed to look for cell cycle

abnormalities in FA cells. One study found that FANCC lymphoblasts had both a spontaneous and MMC-induced increase in the proportion of cells in the G2 fraction (Walsh, et al., 1994b). This defect was corrected by the introduction of the *FANCC* cDNA. In addition, flow cytometry experiments, using patient lymphoblasts, clearly show that the percent of cells in G2 is twice that of controls (Seyschab, et al., 1993). The increase in the G2 fraction can also be directly correlated to DEB sensitivity (Seyschab, et al., 1995). This aberrant G2/M arrest, however, does not represent an abnormal cell cycle checkpoint response, but rather, a normal response to the excessive DNA damage (Heinrich, et al., 1998).

In recent years, the process of programmed cell death, or apoptosis, has become better understood. It is now clear that apoptosis is a well-controlled biological function that is involved in many processes, including embryonic development and hematopoiesis (Cowling and Dexter, 1994, White, et al., 1994). It has also been found that apoptosis and the cell cycle are inter-connected through several proteins, such as p53, that are important in the regulation of both processes. In many cases, p53 is required for DNA damage-dependent cell cycle arrest and can induce apoptosis if the DNA has not been repaired (Bunz, et al., 1998). Given the clinical phenotype of FA and the observed G2 delay in FA cells, it is possible that the defect could be the byproduct of deranged apoptosis. Several labs have looked at both spontaneous and induced levels of apoptosis in lymphoblast cell lines from patients, and have produced contradictory results (Rey, et al., 1994, Rosselli, et al., 1995, Kruyt, et al., 1996, Ridet, et al., 1997). A more recent study, however, has shown that with the introduction of the *FANCC* cDNA, the observed

spontaneous apoptosis can be corrected, as well as the MMC hypersensitivity and cell cycle delay (Buchwald and Moustacchi, 1998). Experiments looking at the involvement of p53 in FA have shown that MMC induces p53 at lower concentrations in FA cells as compared to wild-type. This phenotype can also be corrected with the introduction of the *FANCC* cDNA (Kupfer and D'Andrea, 1996). Additional evidence for the possible role of FANCC in the regulation of apoptosis comes from a study performed on factor-dependent cell lines. In these experiments, the constitutive expression of the *FANCC* cDNA temporarily protects cells from apoptosis caused by factor withdrawal (Cumming, et al., 1996). In addition, hematopoietic progenitor cells from both *FANCC* $^{-}$ patients and *Fancc* knock-out mice, are hypersensitive to interferon-gamma (INF-γ), a molecule that can induce apoptosis (Whitney, et al., 1996, Rathbun, et al., 1997, Wang, et al., 1998). At this time, the majority of the evidence would suggest that there is an apoptotic defect in FA cells. However, the role of the FA proteins in this process remains unclear.

In summary, there are several cellular processes in Fanconi Anemia that appear to be altered, including DNA repair, cytokine production and signaling, cell cycle regulation and apopotosis. Many experiments have been done in an attempt to demonstrate which of these pleiotropic phenotypes are involved in the primary defect of FA. In many cases, however, the results have been inconsistent. Most of these studies were done using immortalized or transformed cell lines from different complementation groups without isogenic controls. It is hoped that with the availability of the cloned genes and knock-out mice, these experiments will yield more conclusive results.

III. The Genetic and Molecular Aspects of Fanconi Anemia

The extensive clinical variability in Fanconi anemia led to the hypothesis that the disease could be genetically heterogeneous. This theory was confirmed by classifying patients into "complementation groups", which are defined by cell-fusion studies that exploit MMC or DEB sensitivity. Hybrids, in which the sensitivity has been corrected (complemented), are assumed to result from the combination of cells from different groups, whereas hybrids that remain sensitive are produced by fusion of cells from the same group. With the use of this complementation assay, eight groups have been identified for FA (FANCA-H) (Duckworth-Rysiecki, et al., 1985, Strathdee, et al., 1992a, Joenje, et al., 1995a, Joenje, 1996, Joenje, et al., 1997). Complementation groups are thought to reflect genetic heterogeneity in that each group potentially relates to a distinct gene playing a part in a biochemical pathway that, when interrupted, leads to a specific aspect of the disease. The recent cloning of the ataxia telangiectasia gene (*ATM*) has, however, challenged this concept. In this case, mutations in a single gene were found in patients that had been assigned to four different complementation groups (Savitsky, et al., 1995). In FA, however, at least four distinct genes (*FANCA*, *FANCC*, *FANCD* and *FANCG*) are known to exist based on their separate chromosomal locations (Buchwald, et al., 1992, Pronk, et al., 1995, Whitney, et al., 1995, Lo Ten Foe, et al., 1996a, Liu, et al., 1997a). Therefore, the "one group = one gene" concept still remains viable for FA, suggesting that at least four genes, when defective, can cause FA. To date, three FA genes have been cloned (*FANCC*, *FANCA*, *FANCG*) and a fourth, *FANCD*, has been

mapped (Buchwald, et al., 1992, Whitney, et al., 1995, Lo Ten Foe, et al., 1996a, de Winter, et al., 1998). Unfortunately, the functions of these genes remain unclear. The prevalence of the FA subtypes has been determined by analyzing patients from both European and US/Canadian origin (Duckworth-Rysiecki, et al., 1985, Strathdee, et al., 1992a, Jakobs, et al., 1997). Approximately 66% of the patients were classified as *FANCA*, ~ 4.3% as *FANCB*, 12.7% as *FANCC*, 4.3% as *FANCD* and the remaining 12.7% are assumed to be *FANCE-H*.

Table 2. Complementation Groups of FA

Subtype	Estimated Percentage of FA Patients	Chromosome Location	Protein Product
A	66.0	16q24.3	163 kD Predicted nuclear localization Shown to interact with FANCC
B	4.3		
C	12.7	9q22.3	63 kD Cytoplasmic and nuclear localization Shown to interact with cdc2, NADPH cytochrome P-450, GRP554 and FANCA
D	4.3	3p25.3	
E	12.7	9p13	65 kD Previously identified as XRCC9
F			
G			
H			

(Adapted with permission from (D'Andrea and Grompe, 1997))

FANCC was the first FA gene to be identified using an expression cloning strategy relying on the hypersensitivity of FA cells to cross-linking agents (Buchwald, et al., 1992). A human lymphoblast cDNA library in the episomal shuttle vector pREP4,

was transfected into an Epstein-Barr virus (EBV)-immortalized FANCC cell line, HCS536. Following selection with both MMC and DEB, complementing cDNAs were isolated from cross-linker-resistant pools of cells and transformed into bacteria for analysis. A set of complementing cDNAs was identified that encoded the same open reading frame. The original HSC536 cell line was shown to have an inactivating L554P mutation on one allele and a 327bp deletion on the second (Gavish, et al., 1993, Parker, et al., 1998). The *FANCC* gene was mapped to chromosome 9q22.3 by fluorescent in situ hybridzation (Strathdee, et al., 1992b). The cDNA consists of 14 exons, ranging in size from 53 to 204 bp, and has an open reading frame of 1,674 bp (Gibson, et al., 1993a). The murine and bovine *FANCC* genes have subsequently been isolated, but no *FANCC* homologues have been reported in lower organisms (Wevrick, et al., 1993, Ching-Ying-Wong, et al., 1997). The murine *Fancc* is able to functionally complement human *FANCC* $^{-}$ patient cells, even though it is only 66% conserved at the protein level.

Mutation analysis of the *FANCC* gene has identified a total of ten pathogenic mutations (Fanconi Anemia Mutation Database http://www.rockefeller.edu/fanconi/mutate). Several of the mutations are common to specific ethnic groups. The most important of these is the IVS4+4A->T splice site mutation, common in the Ashkenazi-Jewish population, which accounts for greater than 80% of FA patients in this group (Whitney, et al., 1993, Whitney, et al., 1994, Verlander, et al., 1995). Patients homozygous for this mutation are more severely affected with multiple congenital abnormalities and early onset of hematological disease. The 322delG mutation is found predominately in patients from Northern European ancestry, particularly from Holland (Verlander, et al., 1994). This

mutation appears to produce a more mild phenotype. Most of the remaining mutations result in either a truncated or internally deleted protein (Gibson, et al., 1993b, Verlander, et al., 1994, Gibson, et al., 1996). Two missense mutations have been identified (L554P and L496R) and have been shown to completely abolish the functional complementing activity of the FANCC protein (Gavish, et al., 1993). Several polymorphisms have also been described (Verlander, et al., 1994).

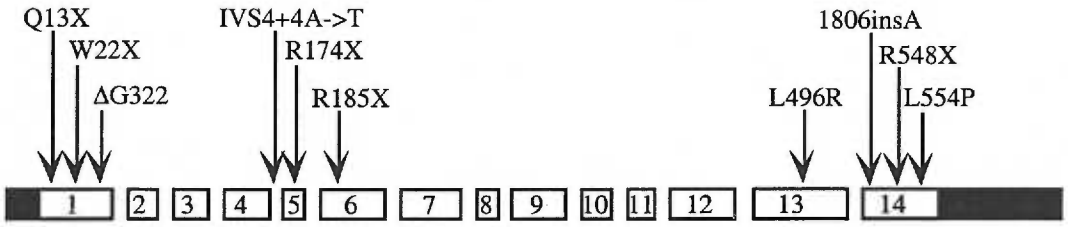


Figure 1. Structure of *FANCC* (Gibson, et al., 1993a). The mutations so far identified in *FANCC* $^{-}$ patients are indicated (Strathdee, et al., 1992b, Gibson, et al., 1993b, Whitney, et al., 1993, Verlander, et al., 1994, Gibson, et al., 1996, Lo Ten Foe, et al., 1996b).

The *FANCC* mRNA is ubiquitously expressed at low levels in adult tissues (Strathdee, et al., 1992b, Wevrick, et al., 1993). There are three different transcript sizes, 2.3, 3.2 and 4.6 kbp, visible on Northern blots. It is believed that these different transcripts are the result of different polyadenylation sites and alternative splicing (Savoia, et al., 1995). RNA in situ hybridization analysis of mouse embryos has shown that *FANCC* mRNA expression is high in the mesenchyme and its derivatives with osteogenic potential (8-10 days post-conception). The *FANCC* transcript is also apparent at later stages of bone development (13-19 days post-conception), and localizes to the cells of the inner perichondrium, periosteum and the zone of endochondral ossification.

Expression is seen in cells from both osteogenic and hematopoietic lineages, and is also detected in non-skeletal tissues (Krasnoshtein and Buchwald, 1996). The *FANCC* mRNA is not induced by MMC and is not differentially expressed during cell-cycle progression (D'Andrea and Grompe, 1997).

The *FANCC* cDNA encodes a 63 kD protein (558 amino acids) that does not show homology to sequences currently compiled in Genbank (Strathdee, et al., 1992b). The protein is predominately localized to the cytoplasm, making a direct role in DNA repair unlikely (Yamashita, et al., 1994, Youssoufian, 1994, Hoatlin, et al., 1998). Forced nuclear localization abolishes the ability of the *FANCC* cDNA to complement cells, therefore supporting that the cytoplasmic localization is essential for the intracellular activity of the protein (Youssoufian, 1996). Immunoprecipitation experiments have identified a set of associated proteins of approximately 34, 50 and 60 kD (Youssoufian, 1995). The level of FANCC protein expression increases slightly during the G2/M transition of the cell cycle. Evidence shows that the FANCC protein binds to the mitotic cyclin-dependent kinase cdc2, suggesting that it may have a cellular function that is regulated during the G2/M transition (Kupfer, et al., 1997b). The interaction of FANCC with cdc2 suggests a molecular mechanism for the observed G2 delay in FA cells. For instance, the FANCC protein may be required for normal cdc2 activity during the G2/M progression. Alternatively, the FANCC protein may be part of a DNA repair pathway downstream of the cdc2 kinase. Yeast two-hybrid screens have also identified several other FANCC interacting proteins, including GRP554 and NADPH cytochrome-P450 reductase (RED) (Hoshino, et al., 1998, Kruyt and

Youssoufian, 1998). GRP554, a molecular chaperone, is localized to the endoplasmic reticulum and is thought to play a role in protein transport. GRP554 is a 94 kD protein and is therefore not assumed to be one of the proteins identified by co-immunoprecipitation experiments (see above). NADPH cytochrome-P450 reductase (RED) is part of a major detoxification pathway and is also involved in the bioactivation of MMC into DNA damaging metabolites. The FANCC/RED interaction requires the amino-terminal region of the FANCC protein and the cytosolic, membrane-proximal domain of RED, and results in the repression of RED protein activity. These results would suggest that, in vivo, FANCC may attenuate the activity of RED, thereby regulating a major detoxification pathway in mammalian cells. Taken together, these interaction studies suggest several different possible roles for the FANCC protein, which may not necessarily be mutually exclusive.

Two murine models of the *FANCC* were developed in which either exon 8 or exon 9 was deleted and replaced by a neomycin resistance gene (Chen, et al., 1996, Whitney, et al., 1996). The cells in both mutants are hypersensitive to cross-linking agents, but the animals themselves do not show any gross morphological aberrations or bone marrow failure. The most predominant feature is a decrease of fertility in both male and female mice. Male mice have reduced testicular weights and a mosaic pattern of normal and abnormal seminiferous tubules. The normal tubules contain all stages of spermatogenesis, including mature sperm, while the abnormal tubules were completely devoid of spermatogenesis. This "all or nothing" pattern is suggestive of germ cell loss. Mutant female mice are unable to carry embryos beyond days 9-10 of gestation. Their

ovaries are reduced in size, hypoplastic and completely lack follicles. Again, these results suggest a defect in germ cell development. Abnormalities involving the reproductive system are also found in FA patients, although less prominent (Alter, et al., 1991, Liu, et al., 1991, Bagnara, et al., 1992).

Despite the fact that mice with the exon 9 deletion had no peripheral blood abnormality, they did have an age-dependent decrease in burst-forming units erythroid (BFU-E) and colony-forming unit GM (CFU-GM) progenitor assays. In addition, hematopoietic progenitor cells showed a distinct hypersensitivity to interferon-γ (IFN-γ). Other mitotic inhibitors had no differential effect. Increased cell susceptibility to INF-γ is mediated by fas-induced apoptosis (Daneshbod-Skibba, et al., 1980). Cells derived from the *Fancc* $^{-}$ mice showed increased levels of fas expression at lower levels of IFN-γ. Based on this evidence, it is possible that INF-γ hypersensitivity is the major pathogenic mechanism in the development of progressive anemia in FA patients.

The *FANCA* gene was identified by two independent approaches, positional cloning and functional complementation (Pronk, et al., 1995, Lo Ten Foe, et al., 1996a). The gene spans 80 kbp, has 43 exons and localizes to chromosome 16q24.3 (Ianzano, et al., 1997). The *FANCA* gene has an open reading frame of 4365 bp and encodes a 1455 amino acid protein. Although the protein has a nuclear localization signal (NLS) and a partial leucine zipper, it does not appear to have any significant homology to any other proteins in the sequence databases, thus supporting the theory that the FA genes are part of a novel pathway.

Over 85 mutations in *FANCA* have been identified, but no predominant alleles are apparent (Fanconi Anemia Mutation Database http://www.rockefeller.edu/fanconi/mutate). The mutations are distributed over the entire gene and include missense, nonsense, splicing, and frameshift mutations. A large number of microdeletions/microinsertions have also been observed and appear to be associated with specific mutation "hot spots" including short direct repeats, homonucleotide tracts, CCTG (CAGG) motifs and TTC repeats. The mutation spectrum of *FANCA* also includes a variety of large genomic deletions. The presence of numerous Alu repeat elements in introns suggests that Alu-mediated recombination might be an important mechanism for the generation of *FANCA* mutations. The presence of such sequence specific hypermutable regions in *FANCA* suggests that the gene may have a higher mutation rate than the other FA genes, which would explain why FANCA accounts for approximately two-thirds of all FA patients (Duckworth-Rysiecki, et al., 1985, Strathdee, et al., 1992a, Jakobs, et al., 1997).

The major *FANCA* mRNA transcript is 5.5 kbp long and appears to be ubiquitously expressed at low levels in the adult tissues tested. In addition to this major transcript, both larger and smaller transcripts can be seen on Northern blots, suggesting a high degree of mRNA splicing and/or alternative poly-adenylation.

The *FANCA* gene codes for a 163 kD protein that has two overlapping bipartite nuclear localization signals at amino acids 18-34 and 19-35. In addition, a partial leucine zipper consensus sequence is found at position 1069-1090. Although the FANCA and FANCC proteins share no sequence similarity, they have been shown to bind and form a complex in the nucleus (Kupfer, et al., 1997a). Little is known regarding the nature of

this interaction. The binding may be a direct protein-protein interaction or it could be an indirect interaction, mediated by other adaptor proteins. Regulated posttranslational modifications of the FANCA or FANCC proteins, such as phosphorylation, may also be required for this interaction. It has been demonstrated that FANCA is phosphorylated in normal and *FANCD*$^{-}$ cell lines but not in FA cells derived from groups A, B, C, E, F, G or H (Yamashita, et al., 1998). Moreover, the phosphorylation can be correlated with both the binding of FANCA to FANCC and the nuclear accumulation of the complex in the nucleus. It is not yet known whether the FANCA/FANCC interaction is constitutive or inducible under various cellular conditions or stress.

The functional importance of FANCA/FANCC protein interaction has been studied by using both point mutations and deletion constructs of the FANCA protein (Kupfer, et al., 1997a, Naf, et al., 1998). Originally, it was observed that the *FANCC* patient-derived L554P mutation prevented the formation of the complex and failed to promote the nuclear accumulation of FANCC. This suggested that the FANCA/FANCC protein binding could be required for normal cellular function. Additional studies showed that mutations or deletion in the amino-terminal nuclear localization signal (NLS) of FANCA resulted in the loss of functional activity, loss of FANCC binding and cytoplasmic retention of the FANCA protein. A DNA construct, in which the endogenous NLS was replaced with a heterologous NLS sequence, resulted in the nuclear accumulation of FANCA but did not functionally correct cells or bind FANCC. Nuclear localization of FANCA is therefore necessary, but not sufficient, for FANCA function.

These results would suggest that FANCA and FANCC perform a concerted function in the cell nucleus required for the maintenance of chromosomal stability.

The cloning of the *FANCG* gene was reported only recently and therefore much less is known about its role in FA (de Winter, et al., 1998). The gene was identified by cDNA complementation in the same manner as *FANCA* and *FANCC* and turned out to be identical to a previously described gene, *XRCC9*. *XRCC9* was identified by its ability to partially correct the hypersensitive phenotype of the Chinese hamster mutant cell line, UV40, to MMC, cisplatin, EMS, UV and gamma-irradiation (Liu, et al., 1997b). Cells from *FANCG* patients are, however, only sensitive to DNA cross-linking agents (de Winter, et al., 1998). The *FANCG* gene has an open reading frame of 1866 bp and encodes a 68 kD protein. The mRNA is also ubiquitously expressed in adult tissues with the highest expression in the testis (Liu, et al., 1997b, de Winter, et al., 1998). Similar to the other FA genes, it too has no homology to any protein in Genbank. Although its function remains unknown, it has been suggested that the *FANCG/XRCC9* gene may play a role in DNA post-replication repair or cell cycle checkpoint control. It is also believed that the FANCG protein may be part of the FANCA/FANCC interacting complex, since it has been shown to interact with FANCA (Alan D'Andrea, personal communication).

IV. The Mapping and Cloning of the *FANCD* Gene: Overview

FANCD was initially localized to a 50 cM region on chromosome 3p22-26 by microcell-mediated chromosome transfer (MMCT) (Whitney, et al., 1995). This technique was employed because of the lack of large families needed for linkage analysis and traditional positional cloning techniques. In addition, MMCT allows for the identification of genes which may be toxic to cells when overexpressed. Several labs have also tried, unsuccessfully, to identify the *FANCD* gene by cDNA complementation. Microcell-mediated chromosome transfer utilizes functional complementation to identify the disease-bearing chromosome.

To identify the chromosome bearing the *FANCD* gene, an immortalized *FANCD*$^-$ patient cell line, PD20, was used as a recipient for chromosome transfer. PD20 cells retain all phenotypic characteristics of FA cells, including MMC and DEB sensitivity, as well as chromosomal instability. Multiple murine A9 cell lines containing a single neomycin-marked human chromosome were used as donors. Microcell-mediated chromosome transfer was used to move the neomycin-marked chromosomes from the A9 cell lines into the PD20 cell line. Microcells were prepared by treating the donor cells with colcemid and cytochalasin B, which inhibit spindle formation and actin assembly respectively, in dividing cells. Because of the inability of the chromosomes to separate, a nuclear membrane forms around as few as a single chromosome. The micronuclei are then physically isolated by centrifugation and fused with the recipient cells using polyethylene gycol (PEG). A microcell hybrid is, therefore, a recipient cell which has

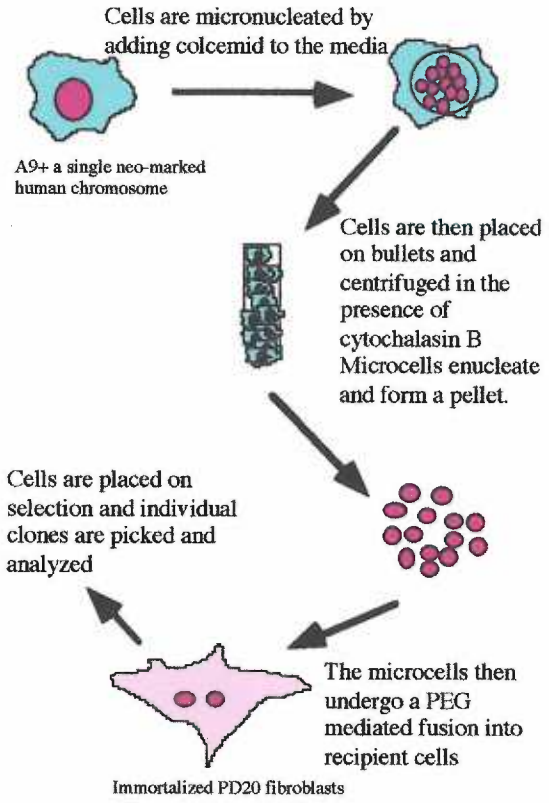


Figure 2. Microcell-mediated chromosome transfer.

taken up a marked chromosome from the donor cell line. Multiple PD20 microcell hybrids containing individual human chromosomes were analyzed. Eventually, it was shown that when human chromosome 3 was introduced into the cells, the sensitivity to DNA cross-linking agents and chromosomal instability was corrected. Deletion analysis of hybrids and exclusion mapping using the FANCD families, localized the gene to a 50 cM region on the short arm of chromosome 3.

Chater 3 describes the original and collaborative work related to the cloning and initial characterization of the FANCD gene. The first section describes the sub-localization of *FANCD* to a 200 kbp region through the generation and analysis of

additional hybrids. Section two discusses the identification of the *FANCD* gene and describes the mutations found in the PD20 cell line. Finally, section three includes the genomic structure of *FANCD* and describes the identification of the murine homolog.

Chapter Two:
Material and Methods

Cell Lines

PD20 is an immortalized FANCD fibroblast cell line generated by the Oregon Health Sciences University (OHSU) Fanconi anemia cell repository (Jakobs, et al., 1996). PD20 lymphocytes were derived from bone marrow samples and were transformed with the Epstein-Barr virus following standard procedures (Doyle, et al., 1993). The mouse cell line, A9, as well as the human/mouse hybrid donor cell line, GM11713 (A9+3), were obtained from the NIGMS Human Genetics Mutant Cell Repository. Murine cell lines including the human/mouse hybrids, were cultured in DMEM with 15% bovine calf serum (Hyclone). Human fibroblasts and microcell hybrids were cultured in α-MEM and 20% fetal calf serum (Summit). Transformed lymphoblasts were cultured in RPMI 1640 supplemented with 15% heat-inactivated fetal calf serum. 2 mM glutamine and 500 μg/ml of penicillin/streptomycin were added to all the media.

Microcell Mediated Chromosome Transfer

Microcell mediated chromosome transfer was performed as described by Fournier (Fournier, 1981). Briefly, donor cells were split onto 150 mm dishes at approximately 40% confluency and allowed to recover for 24 hours. Colcemid (Sigma) (0.06 μg/ml for mouse cell lines and 1.0 μg/ml for human cell lines) was added and cells were incubated for 24-48 hours. Micronucleated cells were then trypsinized and allowed to settle down on "bullets" coated with crosslinked concanavalin A (Sigma). Bullets were then placed into 50 ml centrifuge tubes containing DMEM and 10 μg/ml cytochalasin B (Sigma) and centrifuged at 14,000 rpm (Beckman JA17 rotor) for 30 minutes at 37^{0}C. The resulting

pellets were resuspended in DMEM without serum and filtered through a 5 μm syringe filter (Millex-SV filter unit, Millipore). The microcells were then mixed with the PHA solution (100 ug/ml final concentration phytohemagglutinin P (Difco) in media) and added to a monolayer of recipient cells. After 15 minutes, cells and microcells were fused with 50% PEG for 1 minute (Mercer and Schlegel, 1979), washed with a serum-free medium and allowed to recover overnight without selection. The next day cells were split 1:10 into selective medium containing 400 ug/ml G418 (GibcoBRL). After the selection was complete, clones were picked, expanded and analyzed independently.

Chromosome Breakage Analysis

Chromsome breakage analysis was performed by the Cytogenetics Core Lab at OHSU. For the analysis (Cohen, et al., 1982), cells were plated into T_{25} flasks and allowed to recover overnight. Fibroblast cell lines were then treated with either 40 ng/ml of MMC or 200 ng/ml of DEB for two days. Lymphoblasts were treated with the same concentrations of clastogens, but for three days. After treatment, the cells were exposed to colcemid for 3 hours and harvested using 0.075 M KCl and 3:1 methanol:acetic acid. Slides were stained with Wright's stain and chromosomes were scored for breaks and radials per cell.

Dinucleotide Repeat Analysis

Primers for the microsatellite markers were obtained from either Research Genetics (Huntsville, AL) or Dr. Susan Naylor (University of Texas-San Antonio). PCR

was performed on 50 ng/ul genomic DNA using 0.2 μM of each primer, 0.25 mM dNTPs, 0.15-0.2 mM $MgCl_2$, 1X PCR Buffer (Perkin-Elmer) and 0.25 U Taq DNA polymerase (Perkin-Elmer) for 35 cycles. PCR products were resolved on an 8% denaturing polyacrylamide gel, transferred onto Hybond N membrane (Amersham) and probed with a γ-^{32}P-[ATP] end-labeled 70 base pair CA repeat oligonucleotide. Genomic DNA was prepared as described in (Miller, et al., 1988). For the plus/minus PCR assay, the PCR products were run on 2.5% agarose gels and stained with ethidium bromide.

BACS/P1s

Bacterial artificial chromosomes (BACs) were obtained from Research Genetics after screening the CIT (California Institute of Technology) arrayed human BAC library (Shizuya, et al., 1992, Kim, et al., 1996), using PCR on DNA from hierarchically-pooled BACs. P1 clones were obtained from the library of Pieter de Jong (Ioannou, et al., 1994, Osoegawa, et al., 1998). BAC DNA was purified using a modified protocol from Qiagen. BAC clones were grown overnight in 500 ml LB supplemented with 10 μg/ml thymine, 12.5 μg/ml chloramphenicol. The Qiagen Maxi-prep protocol for single-copy plasmids was followed, with the following changes: 1) volumes of buffers P1, P2, and P3 were each 40 ml; 2) following alkaline lysis, DNA was precipitated with isopropanol, resuspended in 12 ml of QBT buffer containing 50 μg/ml RNase A, and loaded onto the Qiagen Tip-500 column; 3) purified DNA was subsequently dissolved in 300 μl 10 mM Tris, pH 8.0, 0.1 mM EDTA. M13 universal forward and reverse primers were used for end-sequencing of the BACs. BAC end sequences were analyzed using the Blast

algorithm (Altschul, et al., 1990) located on the Baylor College of Medicine Search Launcher website. Repetitive sequences were identified by RepeatMasker (Smit and Green, unpublished). BAC walking was carried out by designing new primers (see appendix A) from non-repetitive BAC end sequence and screening the CIT-arrayed BAC library. P1 clones were purified and analyzed similarly.

DNA Sequencing

DNA sequence determination was performed by the OHSU Microbiology-Core Facility on a model 377 Applied Biosystems (Perkin-Elmer) automated fluorescence sequencer. 500 ng of plasmid DNA was mixed with 3.5 pmoles of primer. For the BACs, 1 μg of DNA was mixed with 12 pmoles of primer. Cycle sequencing was performed with AmpliTaq FS DNA Polymerase using dichlororhodamine dye-labeled terminators, both from Perkin-Elmer .

FISH Analysis

Fluorescence in situ hybridization (FISH) analysis was performed by the Cytogenetics Core Lab at OHSU. For the localization of BACs to 3p25 by metaphase FISH, BAC DNA was nick translation with digoxygenin-11-dUTP (Boehringer-Mannheim) and detected with anti-digoxygenin FITC (Oncor) or rhodamine. The order of the BACs within the contig was determined by interphase FISH (Trask, et al., 1991) using a rhodamine-tagged unique sequence 3p subtelomeric probe (courtesy of David H. Ledbetter) in addition to pairs of labeled BAC probes, one detected with a FITC label

(green), the other detected by a rhodamine label (red). In a second analysis, the subtelomeric locus was labelled with FITC. By comparing the linear order of colored probes, e.g. red-green-red, green-green-red, the order of a given pair of BACs with respect to the telomere was established. A minimum of 50 nuclei were examined for each FISH.

To identify the neomycin-marked chromosome 3 in the microcell hybrids, the eukaryotic expression vector, pXT1 (Stratagene), which contains the neomycin-resistance gene, was used as a FISH probe. After Qiagen purification, the vector was nick translated with digoxygenin-11-dUTP (Boehringer-Mannheim) and detected with anti-digoxygenin FITC (Oncor). The cells were counterstained with propidium iodide and a minimum of 20 metaphases were examined.

I.M.A.G.E cDNA Clones

The following I.M.A.G.E. Consortium (LLNL) clones (Lennon, et al., 1996) corresponding to ESTs in Genbank were purchased from Genome Systems Inc. (St. Louis, MO): clone I.D.s 180467, 1018392, 178261, 198202, 221127, 41789, 31702, 221330, 113501, 82388 and 137340, corresponding to ESTs NIB327, SGC30425, WI-15198, SGC34603, SGC31307, WI-13447, WI-14334, SGC31297, SGC33503, SGC33325 and WI-8719, respectively. The cDNA inserts from these clones were digested with the appropriate restriction enzymes and used as probes for Southern and Northern blot hybridizations.

Southern Hybridizations

Southern hybridization was performed following standard procedures (Southern, 1975). BACs were digested with *HindIII*, resolved on a 0.8% agarose gel and transferred onto Hybond N+ membranes (Amersham). cDNA probes were labeled with random hexamers and α-^{32}P-[dCTP]. Hybridizations were performed at 65^{0}C in 6X SSPE, 0.5% SDS, 5x Denhardts reagent and 100 µg/ml salmon sperm DNA. The membranes were washed for 15' in 2X SSC and 0.1% SDS at room temperature, followed by a wash with 1X SSC and 0.1 % SDS for 15' at 65^{0}C and finally with 0.1X SSC and 0.1% SDS for 5' at 65^{0}C for higher stringency. Inter-Alu PCR products were labeled the same way as the cDNA probes, and then pre-hybridized with 500 µg of human placental DNA. Blots were also pre-hybridized with human placental DNA.

Northern Hybridizations

Human adult and fetal multi-tissue mRNA blots were purchased from Clontech (Palo Alto, CA). Blots were probed with cDNA probes labeled as above and hybridized using the ExpressHyb hybridization solution (Clontech) at 65^{0}C and washed at high stringency, as described above.

Expression Constructs

The *hOGG*1 gene was amplified using primers MG372 5'-ATCGCGGATCCGCGCGGTGCCTGCTGTGGAAATG-3' and MG373

5'-ATCGCGGATCCGCGGAATGGAGGGGAAGGTGCTTAG-3'. The PCR reaction included 0.2 μM of each primer, 0.2 mM dNTPs, 1X Pfu PCR buffer, 1 U *Pfu* DNA polymerase (Stratagene). The conditions were as follows: 35 cycles of 94^{0}C for 45", 55^{0}C for 45" and 72^{0}C for 2'. The PCR products were then digested with *Bam*HI, the sequence of which was included within the primers, and cloned into the pREP4 (Invitrogen) expression vector (map included in figure 3).

A full-length *IRAK-2* cDNA clone was obtained by screening the Strathdee expression library (gift from Dr. M. Buchwald, Hospital for Sick Children, Toronto, Canada) (Strathdee, et al., 1992b). Aliquots of the library were screened by PCR using primers 5'-AGATCATCCTGAACTGGAAACCG-3' and 5'-CTCAGCCTTTCTTACAGAAGCTGC-3' designed from a cDNA sequence in the Genome Database (Accession number AF026273). The following conditions were used: 94^{0}C for 1' followed by 37 cycles of 94^{0}C for 25", 56^{0}C for 25" and 72^{0}C for 25". Once a positive aliquot was identified, serial dilutions were performed until a tube diluted down to less than 10 plasmids, could be identified by PCR. The plasmids from this tube were plated on LB-amp plates and individual clones were analyzed. The positive clones were checked with a second set of primers, 5'-TCTGATTCAAAGGACTTCAGCACC-3' and 5'-AAGACGAATGGCTTCCCGTG-3' with the same conditions as above. A 4.5 kb full-length cDNA was obtained and since the library was made using the pREP4 expression vector (Invitrogen), the clone was directly used for transfection into PD20 cells.

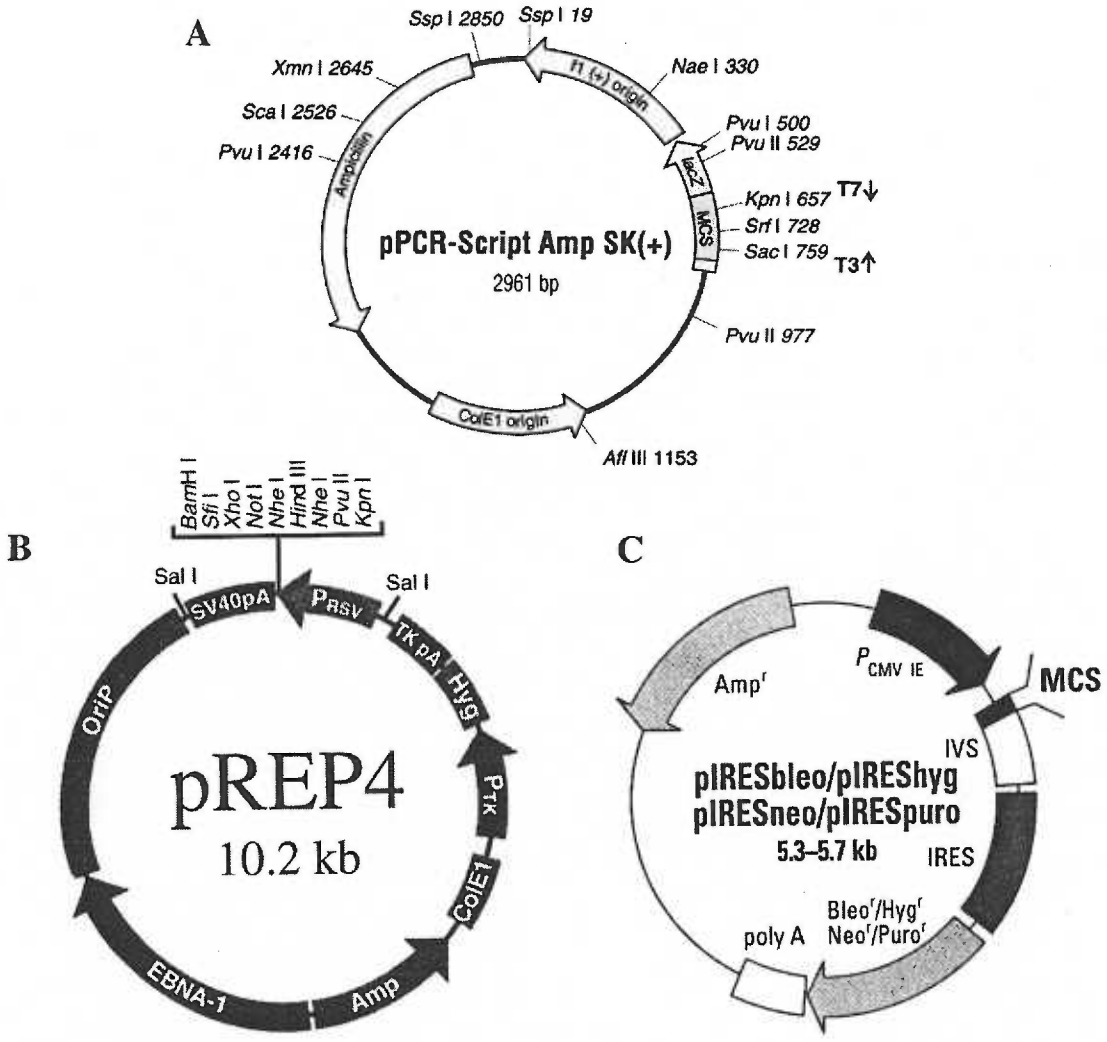


Figure 3: Vector Maps. A) *Pfu* amplified PCR products were cloned into pPCR-Script Amp digested with *Srf*I. B) The *hOGG1* gene and *IRAK-2* were cloned into the mammalian expression vector, pREP4 at the *Bam*HI and *Not*I sites respectively. C) The pIRESneo vector was used to express *SGC34603* in PD20 cells. The gene was cloned into the *Eco*RV and *Not*I sites.

A full-length cDNA for *SGC34603* was constructed by piecing together three PCR products. Figure 4 is a diagram of this construction scheme. The three sections were amplified from a testis cDNA library (Clontech). The PCR conditions were as

follows: 0.2 μM of each primer, 0.2 mM dNTPS, 1X *Pfu* PCR buffer and 1U *Pfu* DNA polymerase (Stratagene), for 35 cycles of 94^0C for 45", 55^0C for 45" and 72^0C for 3'. The PCR products were gel-purified and cloned into pPCR-Script Amp SK(+) (Stratagene) (map included in figure 4). Several clones for each section were completely sequenced to identify clones without PCR errors. Two middle clones were selected so that expression of the L714P polymorphism could be studied. Once a full-length construct was pieced together, several clones were sequenced and two clones, one with a T at position 2141 bp (34603-1) and another with a C at the same position (34603-2), were used for all subsequent constructs. To clone the cDNA into the the pIRES-neo (Clontech) expression construct, the gene was first digested with *Kpn*I, filled in to create blunt ends and then digested with *Not*I. The resulting fragment was then ligated into the *Eco*RV and *Not*I sites. A map of the vector can be found in figure 3.

Electroporation Conditions

All expression constructs were electroporated into the *FANCD* $^-$ patient fibroblast cell line PD20 and a normal control fibroblast cell line, GM639. Cells from an 80% confluent 150 mm plate were harvested, spun down and resuspended in 0.8 ml cytomix (van den Hoff, et al., 1992). 40 μg of purified plasmid DNA was added to the cells in a 0.4-cm-gap cuvette. The cells were chilled on ice for 20' and electroporated at 300 V and 960 μF. The cells were then transferred to two 150 mm plates containing media without selection. 90% of the electroporated cells were put on one plate and 10% on the other.

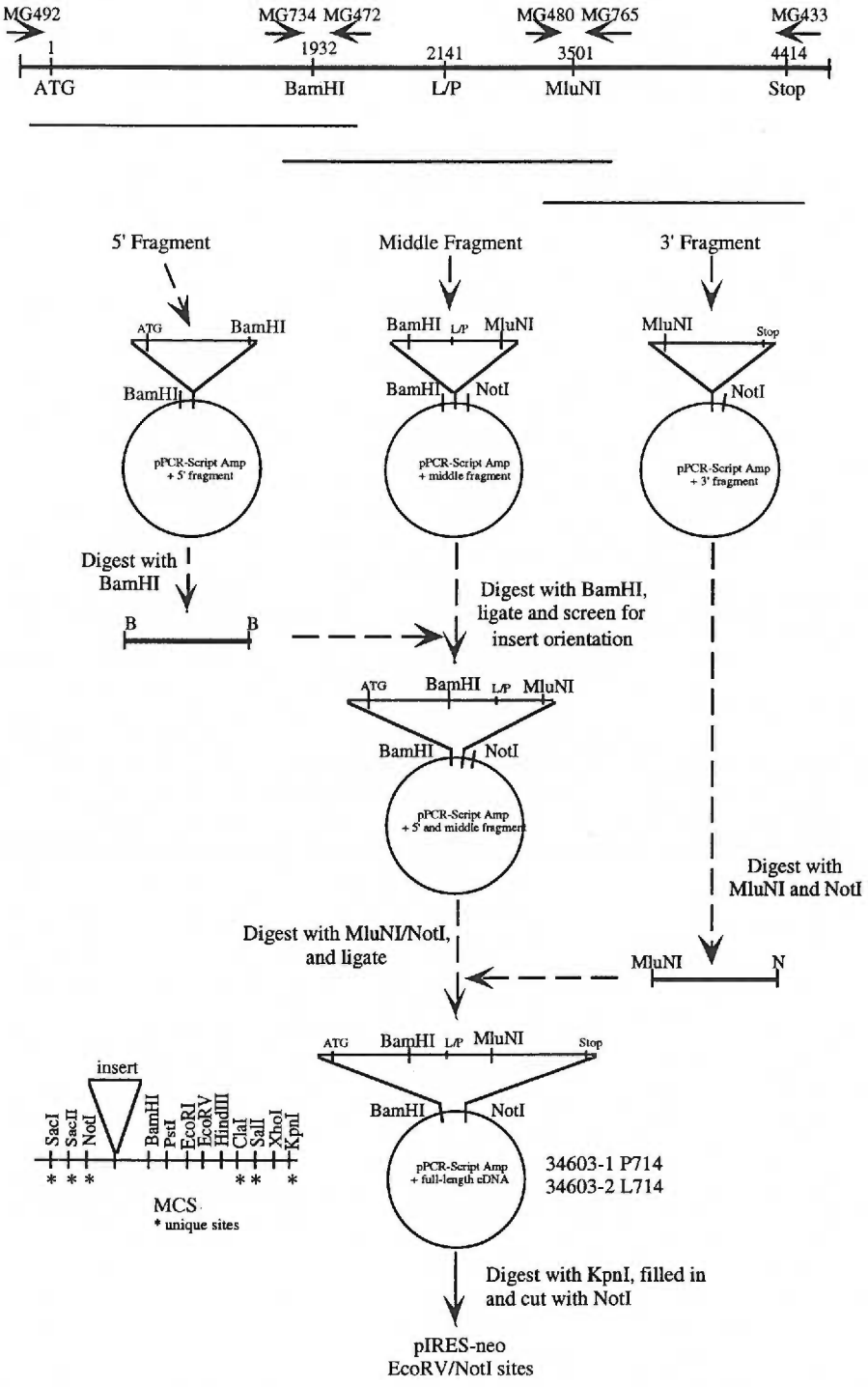


Figure 4. Diagram of cloning strategy for a full-length cDNA of *SGC34603*. Three *Pfu* PCR amplified products were cloned into pPCR-script Amp at the *Srf*I site (this site is destroyed by the insert). The three products were then pieced together by using a *Bam*HI

and *Mlu*NI site. Eventually two full-length clones were obtained, one which codes for a proline at amino acid 714 and another with a leucine. Both cDNAs were then ligated into the pIRES expression vectors (See figure 3 for maps).

After 24 hours the media was changed and the selection drug was added. After selection was achieved, individual clones were picked and analyzed.

Inter-Alu PCR

Three separate PCR reactions were performed on 10 ng of purified BAC DNA using the 3', 5'-ACAGAGCGAGACTCCGTCTC-3', and 5', 5'-GTGAGCCACCGCGCCCGGCC-3' Alu primers. One reaction included both primers and the other two included each primer by itself. Products from all three reactions were pooled and used as probes for Southern hybridizations. The PCR conditions were as follows: 0.2 μM of primers, 0.25 mM dNTPs, 0.2 mM $MgCl_2$, 1X PCR buffer, 0.5 U Taq DNA polymerase (Perkin-Elmer). 30 cycles of 94^0C for 1'15" and 72^0C for 3' were done with a 10' extension at 72^0C at the end of the PCR reaction. The PCR products were combined and purified using a Qiaquick PCR purification column (Qiagen).

Rapid Amplification of cDNA ends (RACE)

5' RACE was performed using a human testis and placenta Marathon-ready cDNA library purchased from Clontech. Reactions were done according to the manufacturer's protocol. Two primers for each cDNA were designed, one for an initial amplification and the second for a nested reaction. Primers MG433 5'-

CTCAGGAGCGATTCTAAGATTGTGTCGTTG-3' and MG 434 5'-GCTGTTATGGAGGGAATGGAAATGGGC-3' were used to amplify SGC34603, primers MG873 5'-ATGCCAATGGAGGGACAGACTTCAGG-3' and MG874 5'-TTCAGGACCAGTCTGGATGGGCTAAGCTAC-3' were used to amplify TIGR-A004X28 and finally primers MG493 5'-GATGTGTGGCAGGGTAACTTGGGACTC-3' and MG494 5'- GAATGGGTGCTTGGAGGAAGGTGTAGG-3' were used to amplify EST- AA609512. The PCR reactions included 1X Klentaq PCR reaction buffer (Clontech), 0.2 mM dNTPs, 0.2 μM primer, 1 U of Advantage Klentaq Polymerase Mix (Clontech) and 2.5 μl of cDNA per 25 μl reaction. The first round of PCR was as follows: 94^{0}C for 1', 5 cycles of 94^{0}C for 30" and 72^{0}C for 4', 5 cycles of 94^{0}C for 30" and 70^{0}C for 4', 25 cycles of 94^{0}C for 30" and 68^{0}C for 4'. The second round was done in the same manner except the last 25 cycles were reduced to 15 cycles.

PCR products were gel-purified and cloned into pT-Adv vector using the AdvanTAge PCR cloning kit (Clontech). DNA from individual clones was purified by standard methods and sequenced using the M13 forward and reverse primers.

The mouse homolog to *FANCD* was amplified from the mouse testis Marathon-Ready cDNA library using primers MG542 5'-CCAGGACAGTGATGAAAGTGACGACAGC-3'and MG543 5'-GCTCACCTCTGACAGACTCTCATCTCTTTGG-3' using the same conditions described above.

RT-PCR Reactions

Total RNA was isolated from cells using the PUREscript RNA isolation kit from Gentra Systems, Inc. (Minneapolis, MN). First strand synthesis was then performed using either an oligo-dT primer or random hexamers. The reactions mixture included 1X first-strand synthesis buffer (GibcoBRL), 0.01M DTT (GibcoBRL), either 10 U of oligo-dT or 50 U of (N), (Parmacia Biotech), 1.25 mM dNTPS, 200 U of M-MLV reverse transcriptase (GibcoBRL) and 1-10 μg of total RNA. The resulting cDNA was used directly in subsequent PCR reactions.

Mutation Analysis of PD20

The full-length *hOGG1* gene was amplified from cDNA made from PD20 lymphoblasts using primers MG373 and MG373 (primer sequence and PCR conditions are listed above). The resulting PCR products were digested with *Bam*HI and ligated into pBluescript II KS (+) (Stratagene). Several individual clones were sequenced using the M13 forward and reverse primers as well as the internal primer MG376 5'-TACCCTGGCTCAACTGTATC-3'.

EST-AA609512 was amplified using primers MG770 5'-TCTCCGTTTACAGTTCAGGACCTC-3' AND MG771 5'-GGTACATACCAAAGGGCATGGTG-3' from genomic DNA prepared from the PD20 fibroblast cell line. Genomic DNA was prepared using standard techniques (Miller, et al., 1988). The PCR conditions were as follows: 0.2 μM of each primer, 0.2 mM dNTPS, 1X *Pfu* PCR buffer and 1U *Pfu* DNA polymerase (Stratagene), for 35 cycles of 94^{0}C for

45", 50^{0}C for 45" and 72^{0}C for 3'. PCR products were gel purified and cloned into the pPCR-Script Amp SK (+) vector. DNA from several clones was sequenced and analyzed.

The 5' end section of FANCD was amplified from its corresponding cDNA after two rounds of PCR. The first round was performed with primers MG471 5'-AATCGAAAACTACGGGCG-3' and MG457 5'-GAGAACACATGAATGAACGC-3'. The PCR product from this round was diluted 1:50 and 5 µl/50 µl reaction was then used for a subsequent round using primers MG492 5'-GGCGACGGCTTCTCGGAAGTAATTTAAG-3' and MG472 5'-AGCGGCAGGAGGTTTATG-3'. The PCR reactions included 1X Klentaq PCR reaction buffer (Clontech), 0.2 mM dNTPs, 0.2 µM primer, 1 U of Advantage Klentaq Polymerase Mix (Clontech) and 1.0 µl/ 25 µl reaction of cDNA. The PCR conditions were as follows: 94^{0}C for 3', 25 cycles of 94^{0}C for 45', 50^{0}C for 45', 72^{0}C for 3' and 5' of 72^{0}C at the end. The PCR products were gel-purified and cloned into the pT-Adv vector (Clontech). DNA from individual clones was purified and sequenced using the M13 forward and reverse primers as well as several internal primers including, MG458 5'-GAAGAAGCCAGTATGGGTG-3', MG750 5'- GATGTCCTTTCAAGCCTCCG-3', MG496 5'- GATGTCCTTTCAAGCCTCCG-3' and MG491 5'-AGAGAGCCAACCTGAGCGATG-3'.

The 3' portion of the gene was amplified as described above but with primers, MG474 5'-TGGCGGCAGACAGAAGTG-3' and MG475 5'-TGGCGGCAGACAGAAGTG-3'. The second round of PCR was performed with

MG491 5'-AGAGAGCCAACCTGAGCGATG-3' and MG476 5'-GTGCCAGACTCTGGTGGG-3'. The 3' end clones were sequenced with M13 forward and reverse primers as well as, MG735 5'-ATGGGATTGCCATAAACCTC-3', MG510 5'-GCGTTCATTCATGTGTTCTC-3', MG447 5'-GCTGGACATTGAGGTCTTC-3', MG480 5'-TGGACCAGGAGTGAAAGTTCAGGAGTACC-3' and MG744 5'-CGGTTTTTTAGATTGCCCAG-3'.

*Msp*I Assay

The mutation in exon 5 changes the A at base pair 376 to a G, thus creating a *Msp*I restriction site. The introduction of this site was used as a diagnostic test to screen 400 chromosomes from the general population. It was also used in the reversion analysis on individual cDNA clones. For genomic DNA, the assay involved amplifying part of exon 4 and all of exon 5 using the primer MG792 5'- AGGAGACACCCTTCCTATCC-3' located in exon 4 and MG803 5'- GAAGTTGGCAAAACAGACTG-3' which is in intron 5. The size of the PCR product was 341 bp which would digest into two fragments of 283 bp and 57 bp if the site was present. The PCR was performed as follows: 0.2 μM of primers, 0.25 mM dNTPs, 0.2 mM $MgCl_2$, 1X PCR buffer, 0.5 U Taq DNA polymerase (Perkin-Elmer) and 50 ng of genomic DNA. 37 cycles of 94^0C for 25", 50^0C for 25" and 72^0C for 35" were done. 5 μl of PCR product was digested in a 20 μl final volume reaction containing 1X L buffer (Roche-Boehringer) and 5U of MspI restriction enzyme (Roche-Boehringer). The digested DNA was then run on a 2.5% NuSieve (FMC Bioproducts) agarose gel.

For analyzing the cDNA clones, PCR was performed using primer MG792 (see above for sequence) and MG753 5'-AGGTTTTGATAATGGCAGGC-3'. The reactions were done exactly as described above with the addition of bacterial cells instead of genomic DNA. In this case the PCR product size was 214 bp which digested into two fragments of 125 bp and 88 bp.

For the L/P polymorphism in exon 23, primers MG733 5'-CAAGTACACTCTGCACTGCC-3' and MG758 5'-TGACTCAACTTCCCCACCAAGAG-3' were used on genomic DNA and primers MG735 5'-ATGGGATTGCCATAAACCTC-3' and MG457 5'-GAGAACACATGAATGAACGC-3' were used on cDNA clones. The assay was done in the same manner as above including the PCR conditions. The genomic fragment was 653 bp which digested into three fragments, 425 bp, 115 bp, and 112 bp when a C is at position 2141 (coding for a proline), or two fragments, 520 bp and 112 bp, when a T is in the same position (coding for a leucine). The cDNA PCR product was 276 bp which digested into either 151 bp, 63 bp and 62 bp when a C was in position 2141 or, 151 bp and 125 bp when a T was present.

Allele Specific Oligonucleotide (ASO) Hybridization

Templates for the ASO assay (Wu, et al., 1989) developed to screen for the exon 37 mutation were generated by amplifying either genomic DNA with primers MG818 5'-AGAGGTAGGGAAGGAAGCTAC-3' and MG813 5'-CCAAAGTCCACTTCTTGAAG-3', or cDNA clones with primers MG452 5'-

AACTACTCAGCCAGAGCGTC-3' and MG830 5'-TGAGGATACTGAAGTCTCGAAC-3'. PCR was performed on either 50 ng of genomic DNA or bacterial cells. The PCR conditions were the same as those used for the *Msp*I analysis described above. The PCR products were denatured in 0.4 M NaOH and 25 mM EDTA and spotted onto Hybond N+ membranes (Amersham) using a dot blot apparatus. Wild-type (5'- TTTCTTCCGTGTGATGA-3') and mutant (5'-TTTCTTCCATGTGATGA-3') oligonucleotides corresponding to the G->A base pair change at position 3707 were end-labeled with γ^{32}P-[ATP] and polynucleotide kinase (Roche-Boehringer). Hybridization conditions were 6X SSPE, 0.5% SDS and 5X Denhardts solution at 40^{0}C for 3 hours. The membranes were washed with 2X SSC and 0.5% SDS once for 20' at room temperature and then for 10' at 42^{0}C.

ChapterThree:
Results

The following people contributed to these results.

Dr. R. Moses Laboratory

James Henja:	Construction of the BAC contig
Donald Bruun:	BAC DNA Preps, Electroporation of IRAK-2
Laura Lucas:	Library screen for IRAK-2

Dr. S. Olson Laboratory

Carol Reifsteck:	Chromosome breakage analysis, FISH
SuEllen Toth-Fejel:	FISH
Nancy Unsworth:	Chromosome breakage analysis
Eleanor Himoe:	Chromosome breakage analysis

Dr. S. Naylor (University of Texas Health Sciences Center, San Antonio, TX)

Dawn Garcia:	Library screen for P1 clones

I. The Sub-localization of *FANCD* to a 200 kb Region on Chromosome 3p25.3

Background. The *FANCD* gene was localized to chromosome 3p22-26 using microcell-mediated chromosome transfer (MMCT) and exclusion mapping in two FANCD families (Whitney, et al., 1995). MMCT was used to transfer individual neo-marked human chromosomes into an immortalized fibroblast cell line from an *FANCD* $^{-}$ patient (PD20). After selection with G418, microcell hybrids were picked and tested for sensitivity to MMC and DEB. In contrast to clones containing other human chromosomes, 6 out of 7 hybrids with chromosome 3 were complemented. The six resistant hybrids were verified by chromosome breakage analysis and found to be cytogenetically corrected, whereas the one sensitive hybrid, PD20-3-8, showed increased chromosome breakage. The results of these studies are listed in Table 3.

Table 3. Chromosome Breakage Analysis

Cell line	DEB (1.2 μM)	DEB (2.4 μM)	MMC (40 ng/ml)	MMC (60ng/ml)
PD20	42	58	24	40
PD20-3-4	0	0	0	0
PD20-3-5	4	4	6	4
PD20-3-6	4	16	4	4
PD20-3-8	48	66	48	44
PD20-3-10	0	2	0	0
PD20-3-11	0	0	0	0
PD20-3-12	0	0	0	2
PD20-3-15	0	4	0	6
PD20-3-20	0	0	0	0
PD20-6914*	0	0	4	18

*Complemented whole cell hybrid between group A (GM6914) and group D (PD20) fibroblast

(Adapted from **(Whitney, et al., 1995)** with permission

Deletion analysis was then used to narrow the region on chromosome 3. This technique can be used to sub-localize genes because chromosomes typically rearrange or are deleted during the transfer procedure. Since both the donor chromosome 3 and the recipient PD20 cells were of human origin, only informative polymorphic markers (i.e. markers in which the donor chromosome 3 and the PD20 chromosomes 3 have different size alleles) could be used. After analyzing a total of 27 hybrids with informative markers, the region of the *FANCD* gene was reduced to the short arm of chromosome 3. Exclusion mapping of two FANCD families further defined the region to 50 cM between markers D3S1307 and D3S1619 (Figure 5).

To narrow the region containing the *FANCD* gene, I chose to focus on the non-complemented hybrid, PD20-3-8. As mentioned above, this hybrid remained sensitive to MMC and DEB even though it appeared to contain an intact chromosome 3p (as determined by the presence of the informative markers). I assumed, therefore, that the donor chromosome 3 had acquired a small deletion that encompassed the *FANCD* gene. By determining the location of this deletion, I could greatly reduce the area in which the*FANCD* gene must reside. Indeed, a deletion in this hybrid was subsequently identified and a BAC contig of the region was constructed. The contig spans approximately 1.2 Mb and contains at least 36 genes. In order to verify the region and possibly narrow it further, four additional non-complemented hybrids were generated and found to have over-lapping deletions of a 200 kb region. This indicates that the *FANCD* gene must reside within this region. The details of the experimental design and results are discussed below.

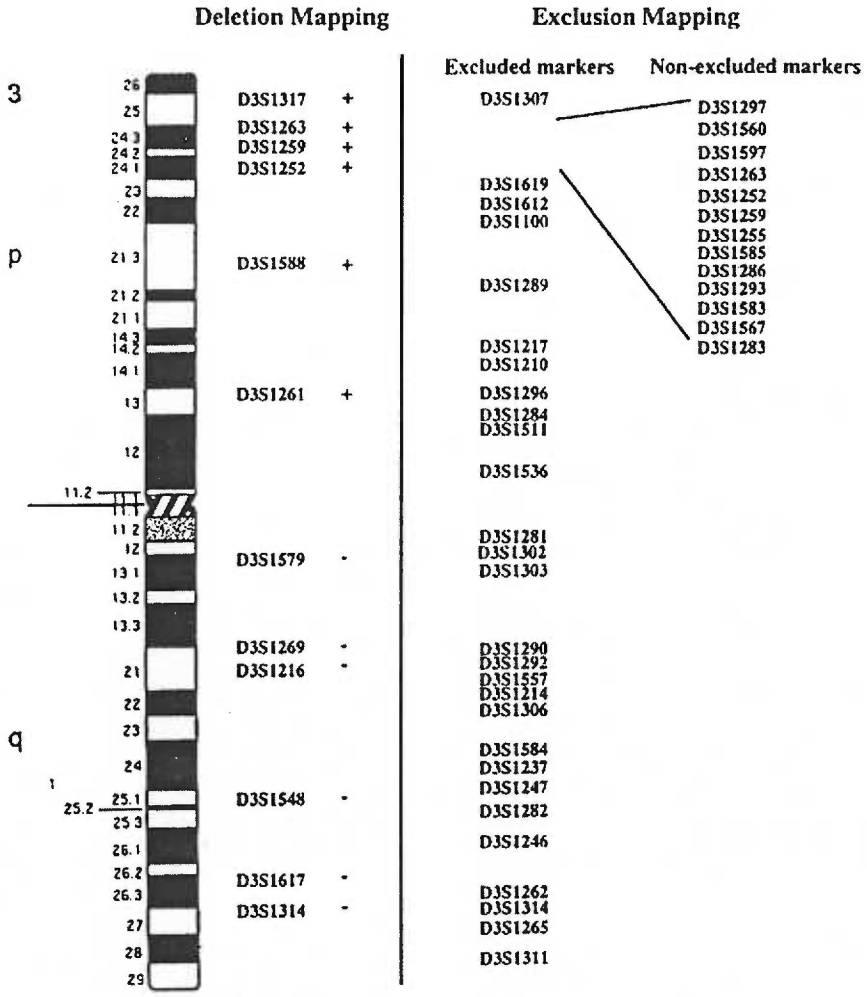


Figure 5. Exclusion and deletion mapping for the *FANCD* locus on chromosome 3. A) Deletion mapping: The 12 informative polymorphic markers are given. The '+' sign indicates markers that were not deleted in any hybrids. 3/24 complemented hybrids (PD20-3-6, PD20-3-15, and PD20-3-20) had deletions of several 3q markers, indicated by a '-' sign. D3S1579 was deleted in one hybrid and D3S1269 and D3S1314 were deleted in two hybrids. B) Exclusion mapping: The 41 markers, which were informative in FANCD family 1, are given. Excluded and non-excluded markers are listed. (Adapted from (Whitney, et al., 1995) with permission)

Characterization of hybrid PD20-3-8. As the above results suggested, the neo-marked chromosome 3 in the PD20-3-8 hybrid contained a small deletion that included the *FANCD* gene. Through the identification of this small deletion, we hoped to further sub-

localize the *FANCD* gene. To identify the deletion, it was first necessary to isolate the neo-marked chromosome 3 away from the other chromosomes 3 in the cell. This was necessary because the density of informative markers on chromosome 3p was insufficient to identify the deletion. As mentioned above, only markers with different size alleles between the donor and recipient chromosomes 3, could be used for this analysis, and this left only 6 markers available to study. By moving the neo-marked chromosome 3 from the PD20-3-8 hybrid into a rodent background, all the genetic markers mapped to chromosome 3p could be used (mouse genomic DNA does not amplify with primers designed for human markers). Figure 6 describes this procedure.

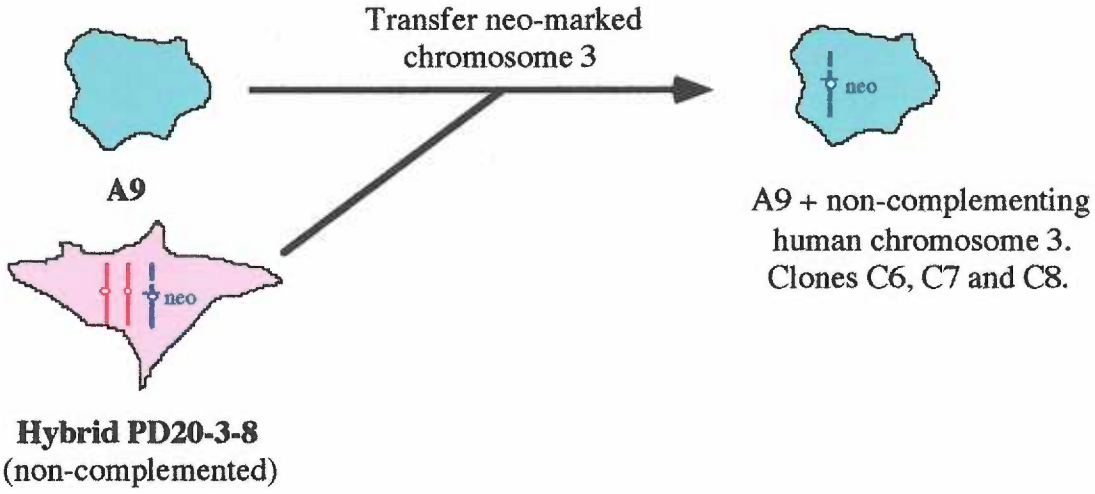


Figure 6. Microcell-mediated chromosome transfer of the neo-marked donor chromosome from hybrid PD20-3-8 into the rodent A9 cell line. Three clones, C6, C7 and C8, were obtained.

The PD20-3-8 hybrid was used as a microcell donor and fused with the A9 (HPRT$^-$) rodent cell line. After selection with G418 (to select for the neo-marked

chromosome 3 from PD20-3-8) and 6-thioguanine (6TG) (to select for the rodent cells), individual clones where analyzed by southern hybridization for the presence of the neo-marked chromosome 3. Three independent clones were identified, C6, C7 and C8. These clones were then used in a plus/minus PCR assay using microsatellite markers spaced approximately every 1 Mb in the previously defined 50 cM region (Whitney, et al., 1995). The +/- PCR assay involved amplifying CA repeats from genomic DNA isolated from the above clones, and running the PCR products on an agarose gel to determine the presence or absence of the marker in the DNA. The A9, A9+3 (the original mouse donor cell line) and PD20 cell lines were used as controls along with the three fusion clones. Using this analysis we identified a single marker, D3S1597, that appeared to be deleted in all three clones (Figure 7).

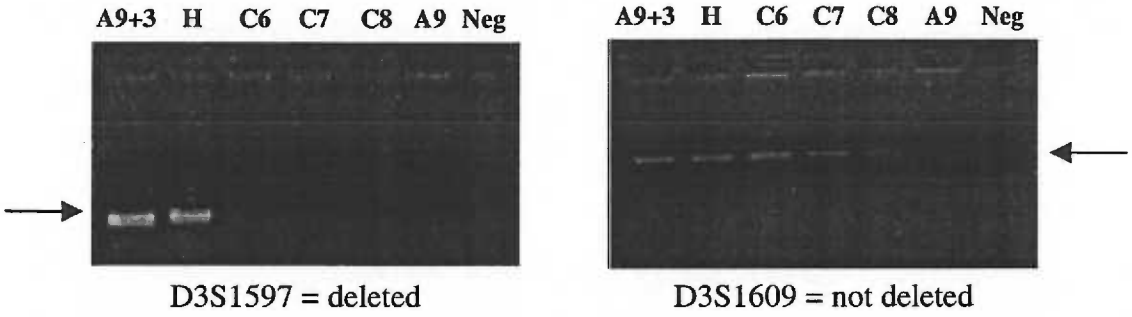


Figure 7. Plus/minus PCR assay to identify deletions in Pd20-3-8 derived hybrids. The marker D3S1597 is deleted in these hybrid. D3S1609 is an example of a marker that is present.

All the available markers, including expressed sequence tags (ESTs), that had been localized around the D3S1597 marker by the CEPH, Genethon and Whitehead Institute maps, were then analyzed (Figure 8). A total of 16 markers, 12 of which were ESTs,

were also found to be deleted in these clones. Based on this analysis, we determined that the breakpoints of the deletion in PD20-3-8 were between markers D3S3591 and D3S3589. Based on the CEPH genetic map of this region, the size of this region was determined to be approximately 8 cM.

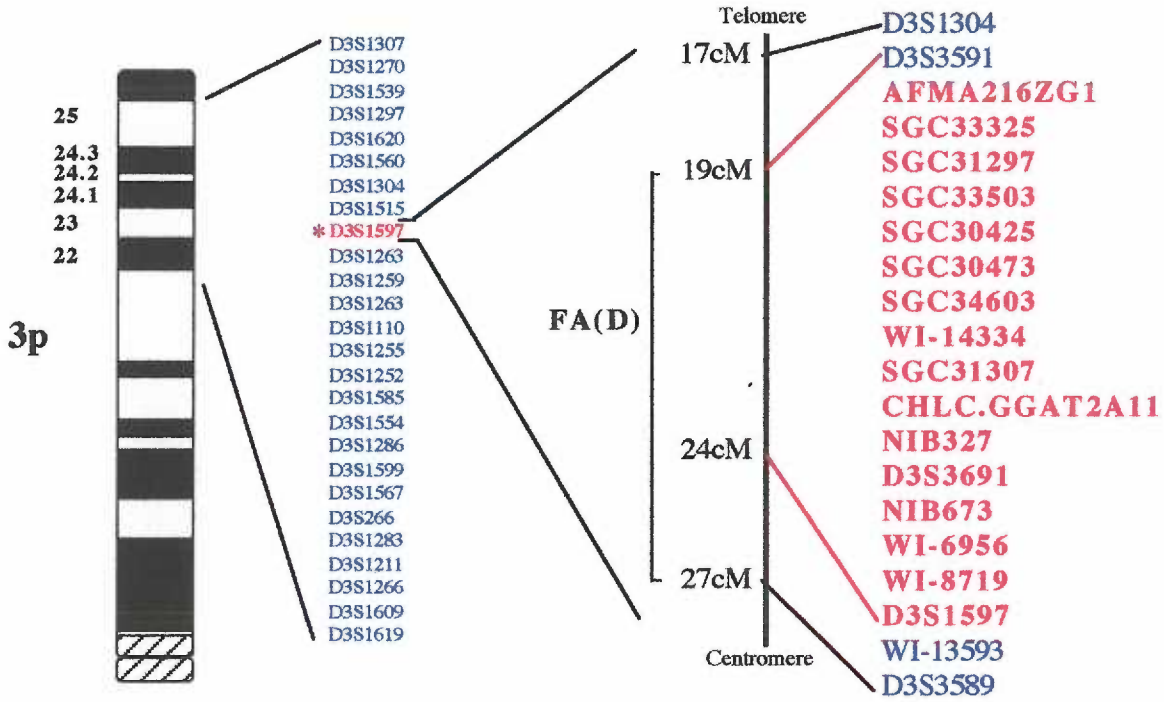


Figure 8. Identification of the deletion in hybrid PD20-3-8. The *FANCD* gene was originally localized to a 50 cM region on chromosome 3p, between markers D3S1307 and D3S1619. Marker D3S1597 (*) was shown to be deleted in hybrid PD20-3-8. Additional markers in this region were then analyzed, narrowing the location of the gene down to a 8 cM region based on a genetic map of the region. Markers in red were found to be deleted in hybrid PD20-3-8.

Construction of a BAC contig of the *FANCD* region. In order to determine the actual size of the newly defined region and the order of the ESTs, it became necessary to build a physical map of the region. Although several yeast artificial chromosomes (YACs) had

been mapped to the region, we decided to build a contig of bacterial artificial chromosomes (BACs), instead of YACs, because of the inherent tendency of YACs to have deletions or to be rearranged. In addition, because of their smaller size, BACs could also be easily electroporated into cells in order to test for complementation.

The 16 STSs (sequence tag sites) and EST markers listed in figure 6, were used to screen the California Institute of Technology (CIT) arrayed human BAC library by PCR (Kim, et al., 1996). BACs carrying these markers were isolated and purified. Localization to chromosome 3 was verified by fluorescent in situ hybridization (FISH) to metaphase spreads of normal human cells. The ends of the human genomic inserts were sequenced using M13 universal forward and reverse primers which flank the pBeloBAC-11 vector multi-cloning site. The Repeat Masker algorithm (Smit and Green, unpublished), available at the Baylor College of Medicine Search Launcher website, was used to identify repeat-free sequences from which additional PCR primer pairs were designed. These PCR primers were then used to search the arrayed BAC library for neighboring BACs. Eventually, small clusters of BACs were formed around the initial set of BACs identified from the 16 markers. As the size of these clusters increased, testing of the newly designed primers on the rest of the contig slowly linked the clusters together. A list of PCR primers developed from BAC end sequence is given in the appendix A.

On two occasions, primers designed from the ends of the BACs failed to identify a neighboring BAC in the CIT library. In the first case, a P1 library was screeened and clones were obtained that connected the two BAC clusters. In the second instance,

however, a gap between two large clusters, flanked by BACs 393J18 and 134A23 (See Figure 10), could not be filled by either BACs or P1 clones. Fortunately, a cDNA clone associated with EST stSG2582, appeared to span the gap. A BLAST search using the 5'-end sequence of this cDNA, produced a high homology score for the end of a BAC in the TIGR BAC sequence database. BAC 2377N12 mapped to chromosome 3p25-26 by FISH, and was shown to overlap with the flanking BACs by PCR analysis.

In order to verify the order of the BACs on the contig, both Southern hybridizations and FISH were used. For Southern blots, BACS were digested with HindIII and probed with cDNA clones that corresponded to the ESTs used to build the contig. Inter-Alu PCR products generated from individual BAC DNA, were also used as probes. For the FISH analysis, the order of two BACs with respect to the p-arm telomere of chromosome 3 was determined by using entire BACs as probes on interphase spreads of human cells. One BAC probe was detected with an FITC-conjugated (green) secondary label while the second BAC probe was detected with a rhodamine-conjugated (red) secondary label. Two chromosomal spreads were then compared, one in which the p-terminus was labeled with FITC and the other in which the p-terminus was labeled with rhodamine. For example, if the linear order on one spread was green-green-red and on the second spread it was red-green-red, then the BAC labeled with FITC would be closer to the p-terminus than the BAC labeled with rhodamine. This allowed the pairwise ordering of BACs at a resolution of 50-100 kb and was crucial for the initial construction of the contig (Trask, et al., 1991). An example of the above study is presented in Figure 9.

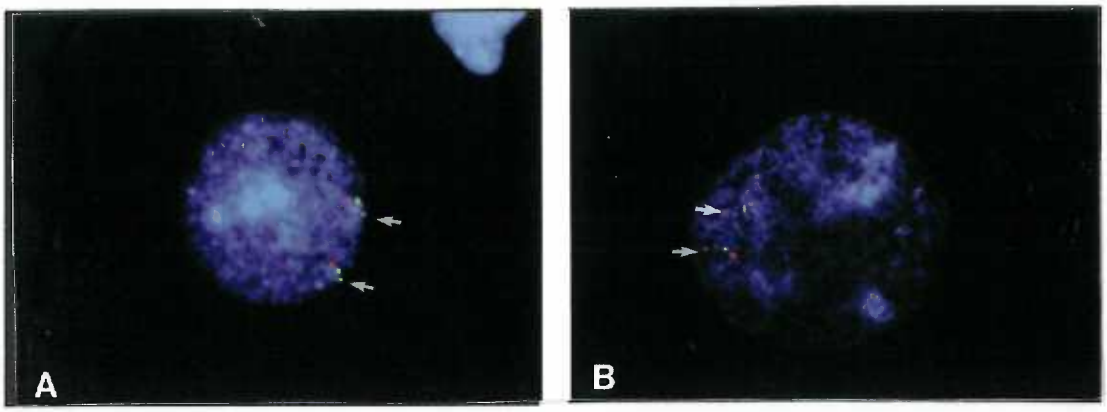


Figure 9. Interphase FISH mapping of two BACs from the contig on a normal human interphase. In order to determine the order of the BACs, two analyses were performed with each BAC pair. The BACs were labeled in the same color, respectively, in both analyses, with the 3p sub-telomere labeled with green in the first and red in the second. **(A)** shows the green sub-telomere probe indicated with the arrow, BAC551O20 in green and BAC168D4 in red. This order was confirmed in the second analysis **(B)** with the telomere probe in red, BAC551O20 in green and BAC168D4 in red.

Localization of additional cDNAs in the BAC contig. The assembled contig is shown in Figure 10. A minimum tiling of overlapping BACs covers 1.2-1.5 Mb. As new ESTs were binned to the 3p25 region by the human genome project, they were directly mapped to specific BACs in the contig by PCR. Assuming one gene per EST, the total number of genes in this region would be at least 36, with a gene density of about one gene per 40 kb. BLAST homology searches to BAC end sequences revealed the presence of the van Hippel-Lindau tumor supressor gene (*VHL*) near the centromeric end of the contig. A 400 kb cosmid contig covering the *VHL* region has been reported, and partially overlaps with the BAC contig described here (Kuzmin, et al., 1994). The order of markers in this overlapping region was in agreement in the two studies.

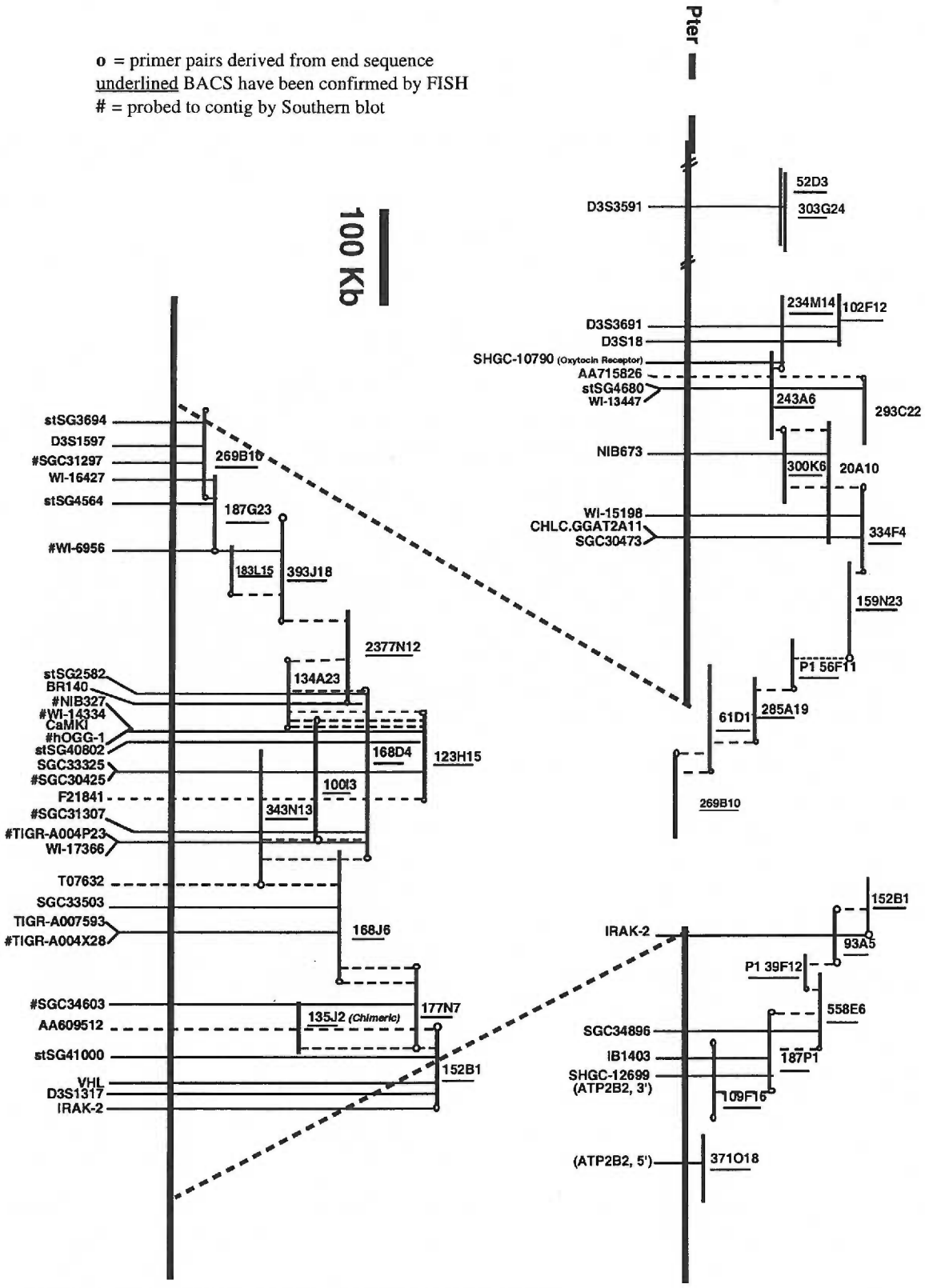


Figure 10: BAC contig on chromosome 3p25.3.

Four additional ESTs that had not been previously mapped to the region, were identified on the basis of BLAST homology searches from the end sequences of several BACs. These include the following ESTs, T07632, AA609512, F21841 and AA715826 (Genbank accession numbers), on BACs 343N13, 152B1, 123H15 and 293C22, respectively.

Because the expression patterns of the three cloned FA genes are similar to each other (ubiquitous expression with highest levels found in liver, testis and thymus), cDNAs mapped to this region were analyzed by Northern hybridization in an attempt to narrow down the number of candidate genes. A summary of the results is given in Table 4. Two genes, NIB673 and WI-6956, appear to be highly expressed in the brain. These genes are of potential interest with regard to two rare syndromes, "3p minus syndrome" and "duplication 3p syndrome" (or trisomy 3p syndrome), in which mental retardation is one of several clinical manifestations (Martin and Steinberg, 1983, Phipps, et al., 1994, Drumheller, et al., 1996). The cDNA corresponding to EST WI-8719 resembled the fibrillin gene and was potentially considered to be a candidate for Marfan syndrome type-2, which has been mapped to this region by linkage analysis (Collod, et al., 1994).

Other well characterized genes that have been placed on the contig include the oxytocin receptor, on BACs 234M14, 61F4 and 243A6, calmodulin kinase-1 (Haribabu, et al., 1995), which was shown to partially overlap with the 3' end of the *hOGG1* gene (Aburatani, et al., 1997), the interferon receptor-associated kinase-2 (IRAK-2) and the human plasma membrane calcium ATPase isoform 2 gene (ATP2B2). Interestingly,

ATP2B2 had recently been associated with mutations causing deafness and imbalance (Kozel, et al., 1998, Street, et al., 1998).

Table 4. Northern Blot Analysis of Mapped Expressed Sequence Tags (ESTs)

EST	BAC	Size	Expression Pattern
NIB673	300K6, 20A10	7.5, 9.5 kb	Highly expressed in brain, low expression in heart, placenta, lung, skeletal muscle and pancreas.
SGC31307	343N13, 100I3, 168D4	3.0 kb	Ubiquitous expression with highest levels in liver, heart, skeletal muscle and pancreas. Low expression in brain.
SGC30425	343N13, 100I3, 168D4, 123H15	6.0 kb	Highly expressed in liver.
SGC34603	135J2, 177N7	5.0 kb	Ubiquitous expression with highest expression in placenta and lowest in skeletal muscle.
WI-6956	187G23, 183L15, 393J18	4.4 kb	Highly expressed in brain and lower in kidney and pancrease.
SGC33503	343N13, 100I3, 168D4, 123H15	1.0*, 1.4, 2.6*, 3.0 kb	Ubiquitous expression with highest levels in heart and skeletal muscle.
EST-AA609512	135J, 177N7, 152B1	1.1 kb	Testis specific.

*These transcripts appear to be present only in heart and skeletal muscle.

Elimination of *hOGG1* and *IRAK-2* as candidate genes for *FANCD*. The human homolog of the yeast gene *ogg* encodes 8-hydroxyguanine glycosylase, an enzyme responsible for the repair of one of the prevalent lesions produced by reactive oxygen species (ROS) [Aburatani, 1997 #1187; Lu, 1997 #1191; Radicella, 1997 #1047; Roldan-

Arjona, 1997 #1193]. Since primary FA cell lines have been reported to be sensitive to oxygen, *hOGG1* was considered as a possible candidate for *FANCD* (Saito, et al., 1993). The full-length normal cDNA was obtained by RT-PCR on mRNA derived from a control cell line. Two splice variants were obtained, one 17 bp longer than the other, and they were both sub-cloned separately into the pREP4 expression vector. In addition, the analogous two cDNA splice variants were amplified from PD20 fibroblasts and also sub-cloned into pREP4. PD20 cells were electroporated separately with each of the expression constructs. After selection with hygromycin B, individual clones were tested cytogenetically for genomic stability after treatment with either MMC or DEB. None of the transfected cells appeared complemented by the addition of the gene. Vector expression of the *hOGG1* was verified by Northern blot. In addition to the expression studies, RT-PCR products from PD20 cells were cloned and sequenced and no sequence changes were observed. Based on these results, the *hOGG1* gene was eliminated as a candidate for the *FANCD* gene.

IRAK-2 was also tested as a possible candidate gene. FA cells have been shown to be hypersensitive to interferon gamma and it is, therefore, conceivable that an IL-1 receptor associated kinase might also interact with several different signal transduction pathways (Rathbun, et al., 1997). A full-length cDNA clone was obtained by screening the Strathdee library (Strathdee, et al., 1992b) (the library was obtained from Dr. M. Buchwald, Childrens Hospital, Toronto, Canada) and was tested for functional complementation of PD20 cells as described for the *hOGG1*. As in the case of *hOGG1*,

there was no indication that the *IRAK-2* cDNA corrected the genomic instability in PD20 cells after treatment with MMC or DEB.

Production and analysis of additional hybrids. In order to confirm and possibly further narrow down the region of interest, additional microcell hybrids were generated with the hope of producing more non-complemented clones that could be used for deletion analysis (as was described for PD20-3-8). Hybrids were generated in the same manner as previously described (Whitney, et al., 1995), and an additional 100 hybrids were analyzed for complementation. Out of the 100 additional clones, seven non-complemented hybrids were identified. The clones were analyzed for deletions using the informative markers D3S3591, D3S1597, D3S1317 and D3S1263. The entire p-arm was deleted in three of the clones, while the other four hybrids (PD20-3-63, PD20-3-71, PD20-3-308 and PD20-3-329) retained the D3S1597 marker. It was, therefore, assumed that these clones had an even smaller deletion than the original PD20-3-8 hybrid. I attempted to move the neo-marked chromosome 3 from these hybrids into the rodent background as had been done with PD20-3-8, but the fusions with these hybrids were unsuccessful. In these experiments neo-resistant clones were obtained, but, when analyzed by PCR, no human chromosome 3 could be identified. Southern blot analysis confirmed this result and demonstrated that the neo-marker appeared to have "jumped" to another chromosome during the transfer process. It is possible that the retrovirus used to mark chromosome 3 with the neomycin gene, was somehow reactivated in the PD20 cell line, which would allow the marker to move to another chromosomal location.

Because we were unable to move the neo-marked chromosome in these clones, we had to use an alternative method to identify their deletions. We chose to use the BACs that had been mapped to the region as probes for FISH on the hybrids. Since these hybrids contained multiple copies of chromosome 3, it was important to be able to differentiate between the neo-marked donor derived chromosome 3 and the endogenous chromosomes 3. This was accomplished by using the neomycin gene as a FISH probe. Once the donor chromosome was identified, the presence or absence of individual BACs from the *FANCD* candidate region was determined. Figure 11 shows a diagram of the deletions found in the additional hybrids. Hybrids PD20-3-63, PD20-3-71, PD20-3-308, PD20-3-329 all had deletions in the same region of the BAC contig. The telomeric break-point in both PD20-3-71 and PD20-3-308 is within BAC 187G23. This region could be a hot spot for chromosome breaks, since this same region was not represented in the CIT BAC library. The deletion in hybrid PD20-3-329 was determined to be within BAC 343N13. The centromeric break-points of all three hybrids were beyond the range of the contig. Hybrid PD20-3-63 was shown to have two deletions, one in the telomeric portion of the contig from BAC 102F12 to BAC 334F4 and a second smaller deletion which spanned a 200 kb approximately the size of BAC 177N7. The deletions in all five non-complemented hybrids overlapped in this one small region. Based on these results, we determined that the *FANCD* gene most likely lies within this 200 kb region were the deletions overlap.

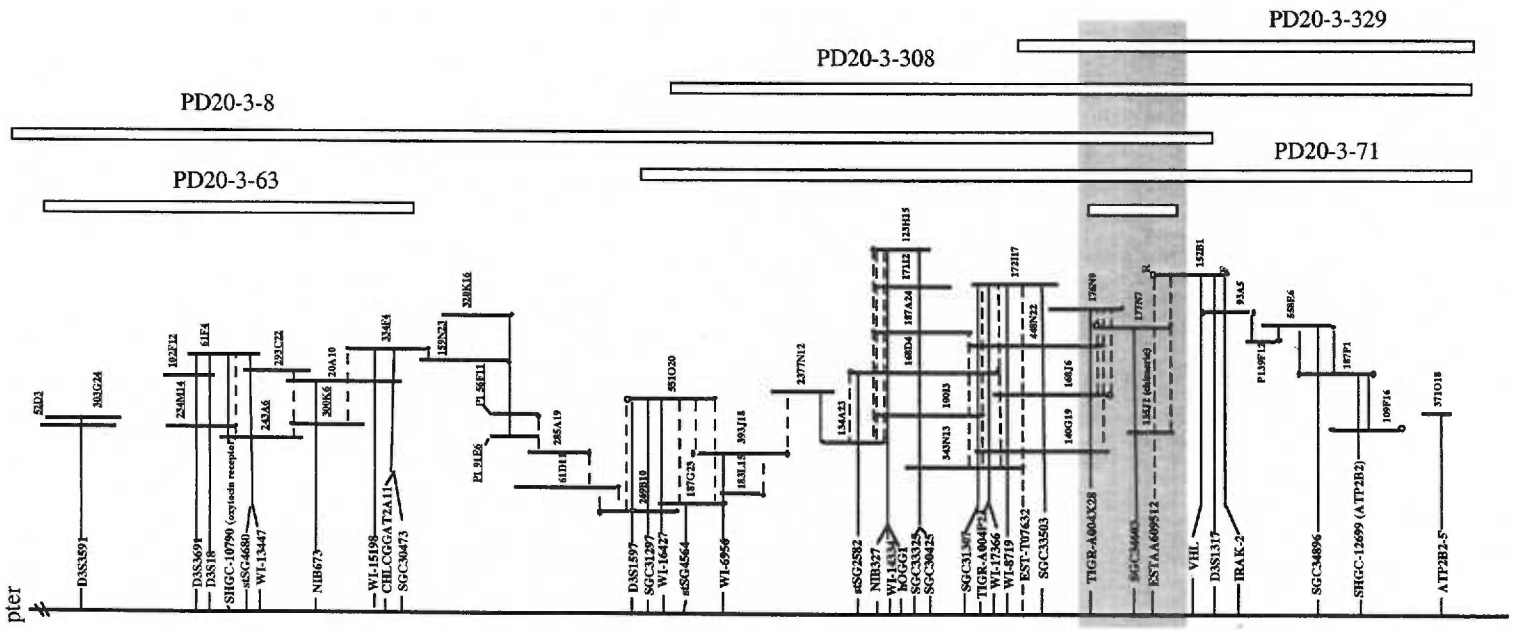


Figure 11. BAC contig with the deletions found in the microcell hybrids. The telomeric breakpoint in the PD20-3-8 deletion is outside the boundary of the contig, but the centomeric breakpoint is within BAC93A5 since D3S1317 is presented but the *VHL* gene is not. Hybrid PD20-3-63 has two deletions in the region, one in top portion of the contig which extend outside the boundary and a second which is approximately 200 kb, encompasses BAC 177N7. The remaining hybrids PD20-3-71, PD20-3-308 and PD20-3-329, all have deletions within the lower half of the contig with their telomeric breakpoint being beyond the contig. The shaded box represents the region of overlapping deletions in the hybrids. This region is approximately 200 kb.

II. The Positional Cloning of the Fanconi Anemia Complementation Group D Gene

Background and Summary: In the previous section, the FANCD gene was sub-localized to a 200 kb region on chromosome 3p25.3 by deletion analysis of microcell hybrids. The hybrids were generated by MMCT of a neo-marked human chromosome 3 into an immortalized FANCD $^-$ patient cell line, PD20. All the hybrids produced were tested for complementation of their cellular sensitivity to DNA cross-linking agents. To identify the location of the gene, hybrids that remained non-complemented after receiving a normal chromosome 3p were analyzed. These hybrids were chosen for analysis based on the assumption that during the transfer process, the donor chromosome 3 had acquired a deletion that encompassed parts of the FANCD gene. By identifying the deletions in these hybrids, the location of the FANCD gene could be elucidated, as long as the deletions were relatively small. Based on polymorphic marker analysis, five hybrids were found to have small deletions on chromosome 3p25.3. The break-points of the deletions in these hybrids were identified and found to overlap in a single region. The region of overlap spanned approximately 200 kb, based on a BAC contig which was assembled in order to facilitate the identification of the deletions as well as to map expressed sequence tags (ESTs) to the region. Three ESTs, TIGR-A004X28, SGC34603 and AA609512, had been mapped to the region of overlapping deletions and, therefore, were considered candidate genes. The full-length sequence of the three cDNAs was obtained and analyzed. Based on expression patterns and homology inquiries, the search was eventually narrowed down to a single candidate gene, SGC34603. This section

describes the analysis of this EST as a candidate for the FANCD gene. The cDNA for SGC34603 was amplified from PD20 lymphocytes and then sequenced. Two pathogenic mutations were found and confirmed in the genomic DNA of the patient. To support the hypothesis that the gene is indeed FANCD, I have shown that a sub-population of the patient's bone marrow has undergone somatic reversion of the mutation. I have also attempted to complement the cellular phenotype, by expressing the full-length cDNA in PD20 fibroblasts. Unfortunately, the results of this experiment suggested that the SGC34603 gene is toxic to cells when over-expressed. It will, therefore, be necessary to use an inducible system to obtain the optimal level of expression. Although cellular complementation could not be shown, we believe that the EST, SGC34603, is the Fanconi Anemia Group D gene based on the mutation and reversion data.

Analysis of three ESTs as candidate genes. The *FANCD* gene was mapped to a 200 kb region on chromosome 3p25.3. By mapping ESTs to a BAC contig constructed in this region, three ESTs were identified that were considered to be candidate genes based on their location. To determine their corresponding transcript size and their expression pattern, multi-tissue Northern blots were probed with PCR products amplified from a testis library using primers designed from their sequence in the Genome Database. Full-length cDNA sequence was obtained for all three ESTs by 5' RACE using the Marathon RACE kit (Clontech). To determine a possible function, BLAST searches with the sequences were then performed to identify any homologies to other proteins. The results of these experiments are summarized in Table 5.

Table 5. Analysis of Candidate Genes

EST	Transcript Size	Northern Blot Expression	Sequence Homology
TIGR-A004X28	800 bp, 1.35 kb and 3.0 kb	Abundant and ubiquitous expression	Mouse fat-specific protein ($P= 10^{-46}$) Cell death activator CIDE-A ($P= 10^{-22}$)
SGC34603	5.0 kb	Low level of ubiquitous expression with highest levels in thymus, placenta and testis.	None
EST-AA609512	1.1 kb	Testis specific	None

The EST, AA609512, has an open reading frame of 900 kb and has only two large exons separated by a small intron. It appears to be testis specific based on its expression pattern on the multi-tissue Northern blots and the fact that it could only be amplified from a testis cDNA library. Because the gene is small and could be amplified from genomic DNA, mutation analysis was performed on PD20 by PCR amplifying the gene from genomic DNA. The PCR products were cloned and sequenced. No sequence changes were identified and EST AA609512, was subsequently eliminated as a possible *FANCD* candidate.

The EST, TIGR-A004X28, is also a small gene with the largest open reading frame identified being only 253 bp. It appears to be highly expressed and is present in all tissues tested. A BLAST search with the cDNA sequence revealed homology to a mouse fat-specific protein and a family of cell death activator proteins (CIDE-A,B). It is worth mentioning that the homology to the cell death activator proteins was only recently identified and the gene was eliminated early on as a candidate based on the homology to the mouse fat-specific protein.

SGC34603 has an open reading frame of 4,413 bp. The start codon is in the second exon, therefore, the first exon contributes to the 68 bp 5' UTR. The gene encodes a 1,472 amino acid protein with a predicted molecular weight of 166 kDa. It is ubiquitously expressed with highest expression in placenta, testis and thymus (Figure 12). This expression pattern matches that of the other known FA genes (Strathdee, et al., 1992b, Lo Ten Foe, et al., 1996a, de Winter, et al., 1998). Based on PCR amplification, the gene appears to have several splice variants. Most of them, however, appear to disrupt the open reading frame and, therefore, are assumed to be non-functional. SGC34603 has no sequence similarities to any human protein in the database which is, again, reminiscent of the other FA genes. Based on these results, SGC34603 was our best candidate for the *FANCD* gene.

Mutation analysis of the SGC34603 gene. To prove that the SGC34603 gene was *FANCD*, we screened the *FANCD*$^-$ patient cell line, PD20, for mutations. RNA was made from both lymphoblast and fibroblast cell lines and reverse transcribed. The cDNA was then amplified in two overlapping sections, 5' and 3', because of the inability to amplify the entire gene in one piece. The PCR products were cloned, sequenced, and analyzed. Two putative mutations were found. The first was an A->G change at base pair 376 in exon 5. In addition to changing an amino acid (S126G), this base pair change may also alter the splice donor site between exon 5 and 6. The base pair change is in the -2 position of the splice donor site which is conserved as an A approximately 60% of the

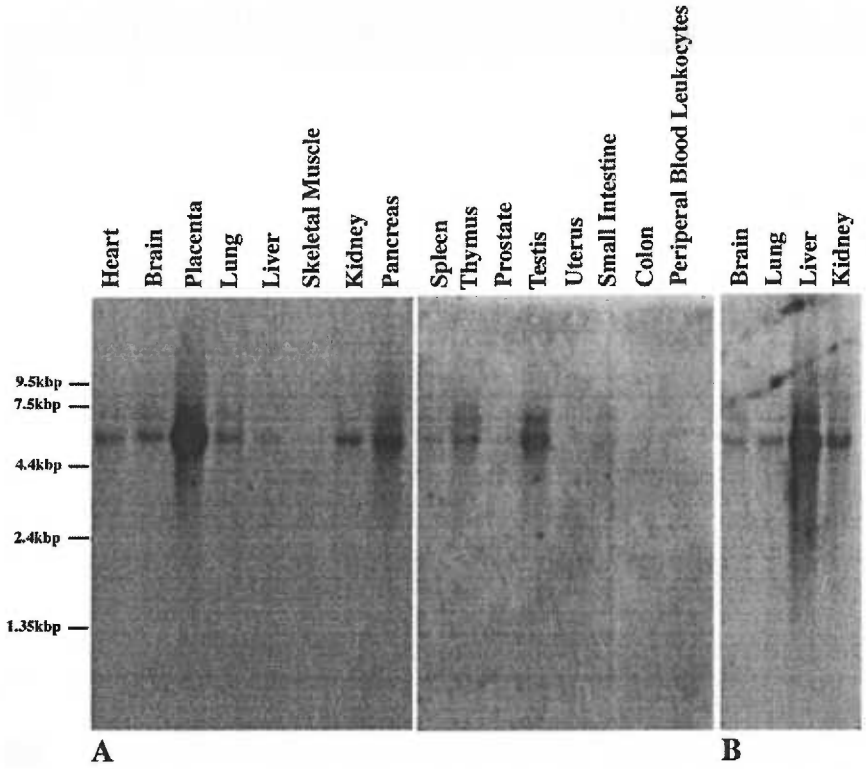


Figure 12: Northern Blot Analysis of SGC34603. Blots were probed with a full-length cDNA and exposed for 24 hours. A) Two adult human blots. B) A fetal human blot.

time and can be a G in some cases (Shapiro and Senapathy, 1987). It is believed, however, that this change has affected this splice site because every clone sequenced with this change had thirteen additional bases of the following intron, inserted into the cDNA. This insertion would disrupt the open reading frame of the gene and produce a severely truncated protein.

The second mutation was somewhat harder to find. After sequencing several 3' clones, 3 clones with exon 37 deleted, were eventually identified. All the remaining clones appeared to be wild-type. Initially it was assumed that the mutation would be in the splice acceptor site since changes to this consensus sequence often results in the

deletion of an exon. However, when the exon was sequenced from PD20 genomic DNA, the only variation found was a G->A base pair change at position 3707, in the middle of the exon. This change creates an arginine to histine amino acid substitution. Even though this is a substantial amino acid change, it is also possible that the change may be affecting the stability of the mRNA, since the allele appeared to be underrepresented in the cDNA population and the only clones that were identified had exon 37 deleted.

Both mutations have been shown to also be present in the genomic DNA of the patient. The mutations were found to be on different alleles by sequencing the genomic DNA of the parents (Figure 13). In order to show that these mutations are not common population, by screening a panel of 200 genomic DNA samples. The exon 5 mutation creates a MspI site and could, therefore, be screened for with a restriction digest. An allele specific oligonucleotide (ASO) assay was used for the exon 37 mutation. Neither one of the putative mutations was found in the general population. The family history of the patient was also checked for any specific ethnic background that would not be represented in the screened population. The family is of Northern European decent which corresponds with the DNA samples used in the polymorphism screen.

PD20 is heterozygous for three additional base pair changes, 1122G->A, 2141C->T and 4098G->T. All but one were conservative changes and were, therefore, assumed to be polymorphisms. One base change converted a leucine to proline in exon 23 at amino acid 714. Typically this might be considered a pathogenic substitution. However, the allele is present in approximatley 50% of the general population, and, therefore, represents a polymorphism. The position of these variants is shown in Figure 14.

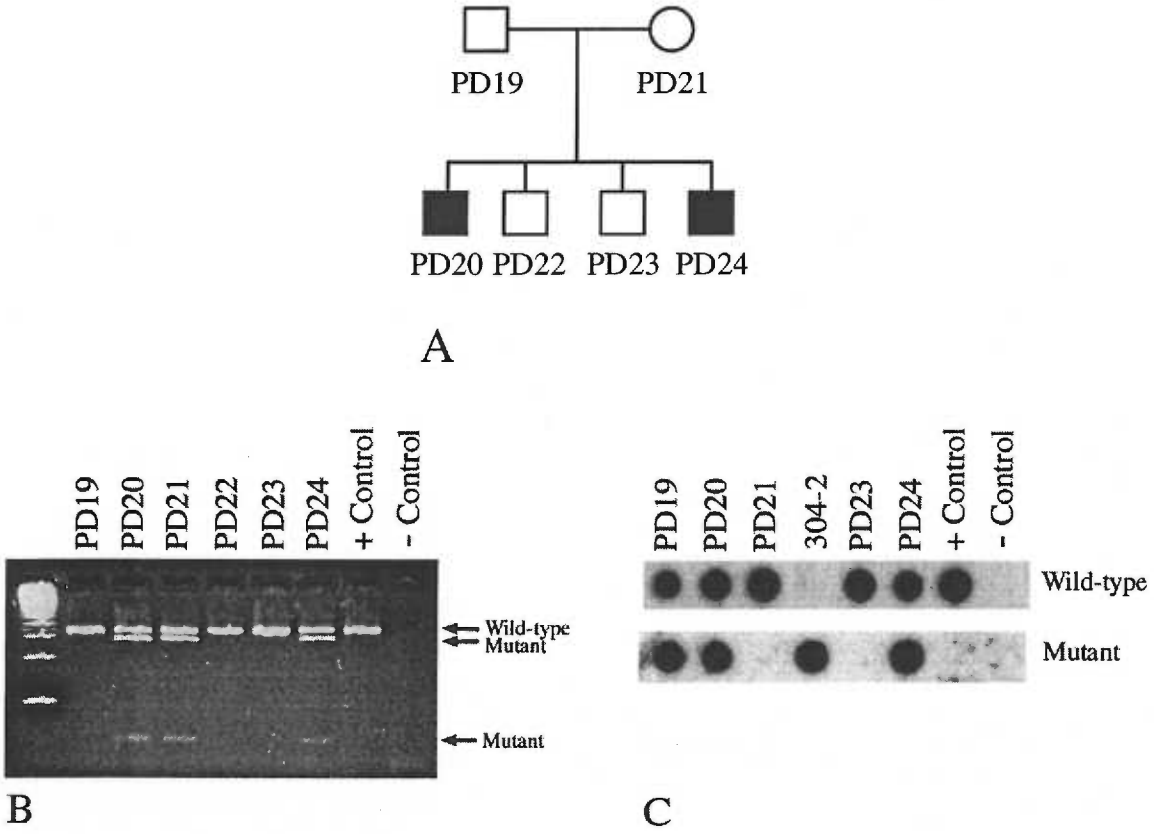


Figure 13: Mutation analysis of the PD20 family. **A)** Pedigree of the PD20 family. **B)** The mutation in exon 5 creates a MspI site, therefore, the site can be used as a marker for carrier status of the mutation. The exon 5 mutation in PD20 is on the maternal allele. **C)** An ASO assay was developed to screen for the exon 37 mutation. The DNA sample, 304-2, represents a rodent hybrid that contains the paternal allele from the PD20 cell line. This figure shows that the exon 37 mutation is on the paternal allele.
polymorphisms, the presence of the sequence changes was investigated in the general

Reversion analysis of PD20 lymphocytes. Recently, a bone marrow sample from the PD20 patient was received. Since the patient's blood counts had begun to rise, it is possible that a somatic reversion of his mutations may have occurred in a sub-population of the bone marrow. Approximately 25% of FA patients show evidence of

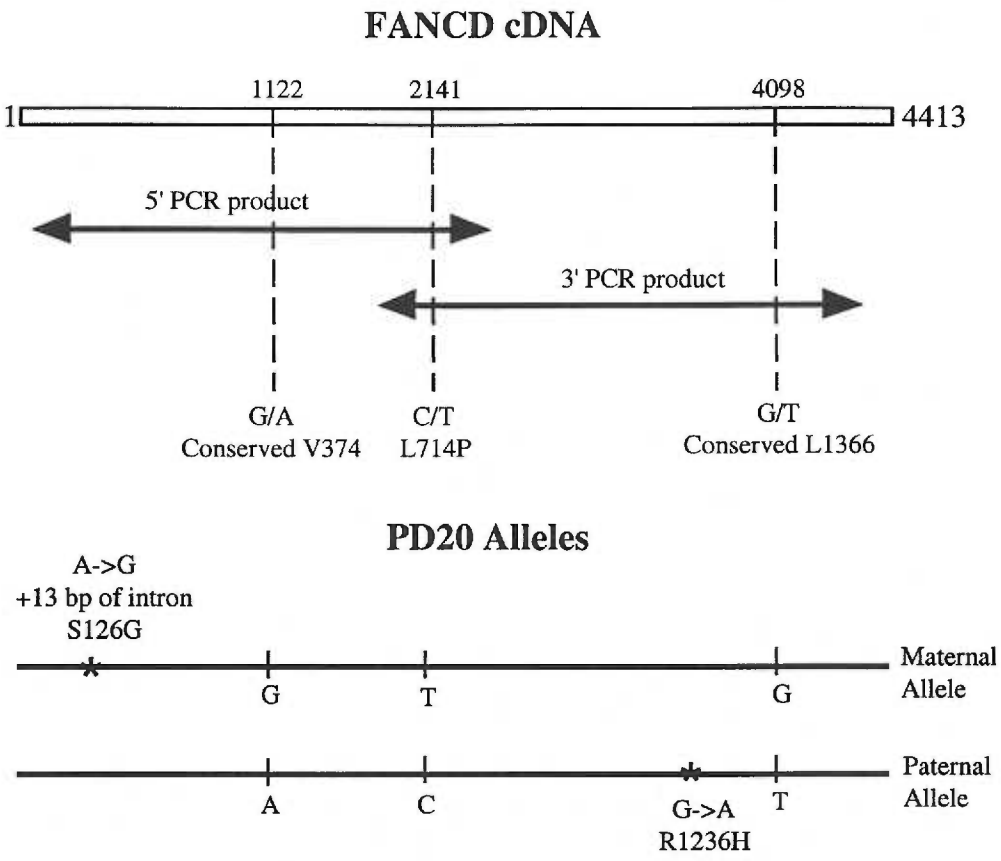


Figure 14. Mutation Analysis of *FANCD* in PD20. The *FANCD* cDNA is shown at the top. The gene was amplified from PD20 cDNA in two sections, 5' and 3'. PCR products were cloned and sequenced. Two mutations were found. The first was an A->G base pair change in exon 5. The clones with mutation also had 13 additional bases of intron sequence inserted. The second mutation was a G->A base pair change in exon 37 causing a R->H amino acid substitution. PD20 was found to be heterozygous for three polymorphisms.

spontaneously-occurring mosaicism as manifested by the presence of two populations of lymphocytes, one which is hypersensitive to cross-linking agents while the other behaves normally in response to these agents (Lo Ten Foe, et al., 1997). It has been shown that this type of mosaicism is most often the result of an intragenic mitotic recombination, which would result in the production of one wild-type allele and one

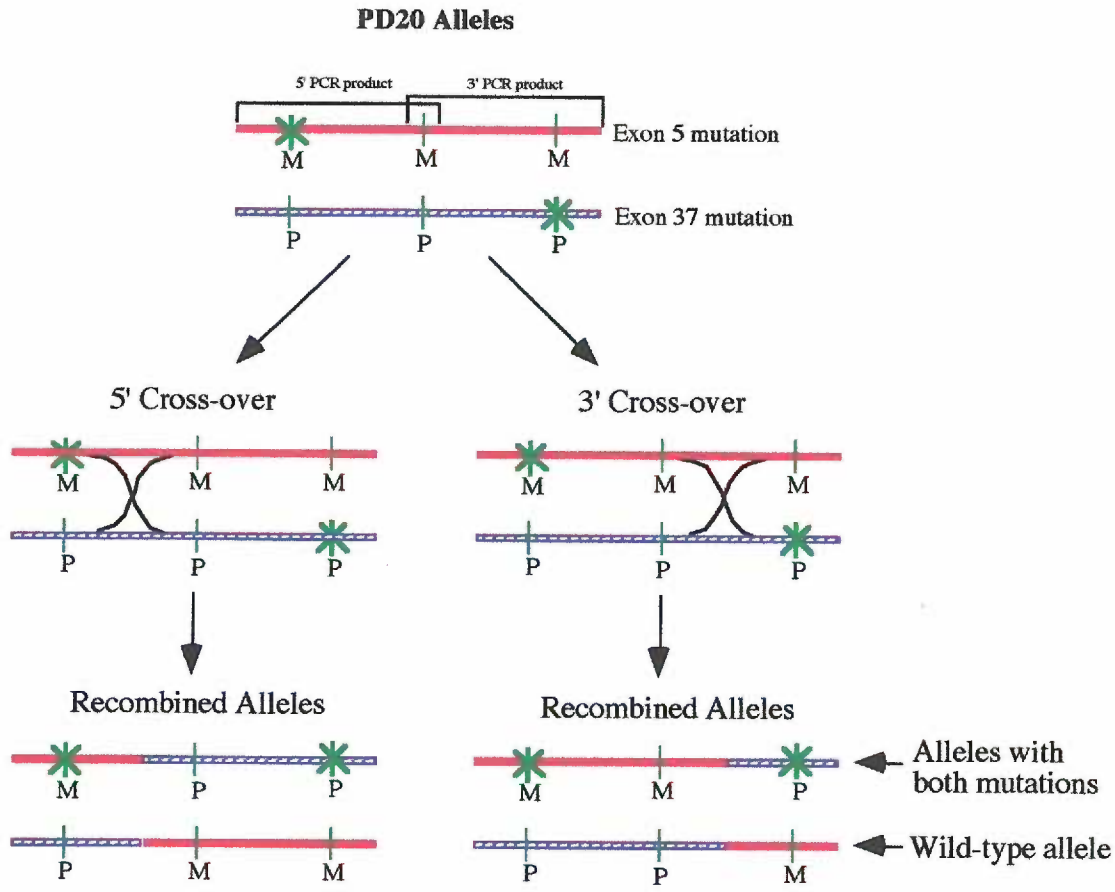
allele with two mutations. If the reversion occurred in a hematopoitic progenitor cell, it would give rise to several lineages of phenotypically-corrected cells. It is thought that these cells would have a selective advantage over the non-reverted cells because of their ability to produce a wild-type protein and would eventually repopulate the bone marrow. Since PD20 appears to be a compound heterozygote and the mutations are at opposite ends of the gene, it would seem that the chance of this type of event occurring would be fairly high. This would also explain the rise in the patient's blood counts. If this were the case, it would also further support the hypothesis that SGC34603 is *FANCD* since there would be no selective advantage for this type of event in a random gene. Notably, it has never been shown that correction of the phenotype was the result of an acquired mutation in another gene.

To demonstrate that a recombinational event was responsible for the phenotypic correction of the cells, a correlation between the phenotype and genotype of the lymphoblasts at the *FANCD* locus, was made. Lymphocytes from the bone marrow sample were transformed with EBV. Chromosome breakage analysis on the population of cells showed a decreased sensitivity to MMC and DEB (15-20% radials) as compared to previous PD20 lymphoblast cells (40% radials). This suggested that, both complemented and non-complemented cells existed in the lymphoblast population. RNA was made from the cells and reverse-transcribed. The gene was again amplified in two overlapping sections (5' and 3'), as had been done for the mutation analysis (See Figure 14). Two separate PCR reactions were performed for each section to eliminate the possibility that the event was due to template switching during PCR. All four products

were cloned and transformed separately. Forty-six colonies were picked from each transformation and analyzed. The phase of the alleles was previously determined, by analyzing DNA from the parents of PD20 (See Figure 13). Because PD20 is heterozygous for a polymorphism in the middle of the gene (L/P714) which was present on both PCR products, cross-overs between the mutations and the polymorphism could be identified.

To identify a cross-over in the 5' portion of the gene, the region around exon 5 (maternal mutation) and exon 23 (L/P714 polymorphism) was amplified directly from the 92 bacterial colonies (46 from each transformation). To determine the phase of each colony, the PCR products were digested with *Msp*I. The exon 5 mutation creates an *Msp*I site and the L/P714 polymorphism deletes an *Msp*I site. The products were resolved on an agarose gel and their analysis revealed that a cross-over had occurred between the exon 5 and the polymorphism. The results of this experiment are summarized in Figure 15. 44% of the clones had the paternal exon 5 allele and the maternal exon 23 polymorphism, the reverse was seen in 1.3% of the clones. The remaining 54.7% had the paternal exon 5 allele and paternal polymophism. The maternal allele was not observed. Since the cross-over was seen in clones from both PCR reactions, template switching during the PCR is an unlikely explanation.

The 3' end clones were also analyzed. For these clones, the region surrounding exon 23 (L/P714 polymorphism) and exon 37 (paternal mutation) was again amplified directly from the 92 bacterial colonies transformed with the 3' PCR products. The exon



A

5' PCR Products			
M M	P P	M P	P M
0%	54.7%	1.3%	44%
3' PCR Products			
M M	P P	M P	P M
89%	0%	0%	19%

B

Figure 15: Reversion Analysis in PD20. **A)** An intragenic recombination event between the maternal (M) and paternal (P) alleles would result in one allele without mutations and the other with two mutations. This cross-over could ocurr between either exon 5 (maternal mutation) and the polymorphism or the polymorphism and exon 37 (paternal mutation), resulting in the alleles shown. **B)** To identify the presence of such an event in

the PD20 lymphoblasts, RNA was made and reverse transcribed. The SGC34603 gene was then amplified from cDNA in two sections, 5' and 3'. The PCR products were cloned and 92 individual bacterial colonies were analyzed to determine which allele they contained. As is shown in the table, almost half of the 5' end clones had reverted alleles while 19% of the 3' end clones did.

23 products were digested with *Msp*I as above, while an ASO assay was used to analyze the PCR products from exon 37. Again a cross-over between the polymorphism and the exon 37 mutation was observed. In these clones, 19% retained the paternal allele at polymorphic loci and the maternal allele exon 37. The remaining 81% were composed of the maternal allele for both loci. Neither the paternal allele nor the complement to the cross-over were observed.

Since only half of the gene could be analyzed at a time, it is possible that the paternal 5' end clones (54.7%) actually represented a cross-over event which occurred in the 3' portion of the gene. The same could be true for the 81% of the 3' end clones that retained the maternal. If this were the case, the number of cells without a cross-over would be very low.

The above analysis supports the hypothesis that a sub-population of cells in the bone marrow of PD20 have undergone a somatic reversion. Although we did not expect to find evidence for two different cross-over events, it is possible that either both events occurred in vivo or that one is the product of the EBV transformation. Analysis of the reverted clones, strengthened the theory that SGC34603 is the *FANCD* gene.

Expression studies with SGC34603. Another proof of identity for a candidate FA gene is complementation of the cellular defect by expressing the gene in a patient cell line. A full-length cDNA was constructed by piecing together three PCR products with restriction enzymes. Each piece was amplified from the Marathon testis cDNA library (Clontech), using Pfu polymerase (Stratagene). The PCR products were cloned into pPCRscript-amp (Stratagene) and several clones for each section were sequenced to check for splice variants and PCR errors. The three sections were then pieced together using a *Bam*HI and *Mlu*NI site at 1932 bp and 3501 bp, respectively (refer to Figure 4). Two separate constructs were made, one which codes for a leucine at position 714 and one which codes for a proline at the same position, to make sure that this polymorphism had no effect on expression. Both constructs were cloned into the bicistronic vector, pIRES-neo (Clontech) (refer to Figure 3). This type of vector insures expression of the gene of interest under selection, because a single transcript is made of the gene of interest and the selectable marker. If the gene is interrupted during integration, the cells will no longer be G418-resistant, and only cells expressing the gene of interest will grow under selection. The two constructs were electroporated into PD20 fibroblasts along with a construct containing the GFP gene as a control. After two weeks of selection with G418, resistant clones were present on the GFP control plate, but not on the plates with SGC34603. We repeated this experiment, and the same result was obtained. To verify this result, we did a mixing experiment in which we transfected the cells with a 50/50 mixture of the SGC34603 construct and the GFP construct. All the cells that survived showed fluorescence. From these experiments, we concluded that when the SGC34603

gene is overexpressed, it is toxic to the cells. It will, therefore, be necessary to express the gene in an inducible system where the expression can be tightly regulated.

III. The Genomic Structure of the *FANCD* Gene and the Identification of a Murine Homolog

Background and Summary. As described in the previous section, the Fanconi anemia complementation group D gene (*FANCD*) was identified by a positional cloning method. The gene has an open reading frame of 4,413 bp encoding a 1,472 amino acid protein and is localized to chromosome 3p25.3. FANCD has no significant homology to any known human proteins, including the other cloned FA genes. The gene appears to have many naturally occurring splice variants, which will potentially make it difficult to search for mutations in the cDNA of patients. It will, therefore, be easier to screen for mutations in the genomic DNA of patients, which is readily available from blood samples. Accordingly, knowledge of the genomic structure will be necessary. In pursuit of these means, a bacterial artificial chromosome (BAC) has been identified which contains the entire coding sequence of the gene. The cDNA sequence of the gene was used to design primers that were subsequently used to directly sequencing the BAC DNA. The intron/exon junctions were identified and primers were designed from intronic sequence to amplify all the exons.

In addition, the murine homolog for *FANCD* gene was identified. The full-length sequence was obtained and compared to the human gene. Knowledge of the cDNA sequence will aid in the production of a knock-out mouse.

Genomic organization. During the construction of the BAC contig for the *FANCD* gene region, two BACs, 177N7and 135J2, were identified in a screen using primers designed

for the 3' end of SGC34603 (refer to Figure 10). To determine the direction of the gene with respect to the contig, all of the BACs in the region were digested and probed by Southern blot with the full-length *FANCD* gene. The results revealed that the 5' end of the gene is telomeric to the 3' end. It also appeared from this blot, that the entire gene could be present on BAC 177N7. Primers designed from both ends of the gene confirmed this result. DNA from this BAC was, therefore, used for all subsequent sequencing reactions. The genomic size of the *FANCD* gene was also determined to be approximately 85 kb, as estimated from restriction analysis and Southern blot. (Figure 16).

The intron/exon junctions and at least 60 bp of additional intronic sequence were obtained by directly sequencing BAC 177N7 DNA, with primers designed from the *FANCD* cDNA sequence. Sequence analysis confirmed the 2141C/T variant as a polymorphism, which causes a leucine to proline amino acid substitution (L/P714), and identified an additional conserved polymorphism, 1509C/T. The gene consists of 43 exons ranging in size between 35 to 148 bp (Table 6). Splice junction boundaries agreed with published consensus sequences (Shapiro and Senapathy, 1987) for donor and acceptor splicing sites. The splice donor sites all start with "GT" and over half of the splice acceptor sites end in "CAG" (54%), the remainder, end with either "TAG" (36%) or"AAG" (10%). The splice sites were scored according to Shapiro and Senapathy (Shapiro and Senapathy, 1987) and the scores for both the donor sites and acceptor sites ranged from 73-93. PCR primers were designed in the intronic sequence surrounding the exons and tested for the amplification of the 43 exons from genomic DNA. Introns 4, 5,

8, 11, 12, 13, 21, 22, 24, 27, 31 and 35, were small (Table 6) so the corresponding exons were combined in some of these cases. The primer sequence and PCR product sizes are listed in Table 7.

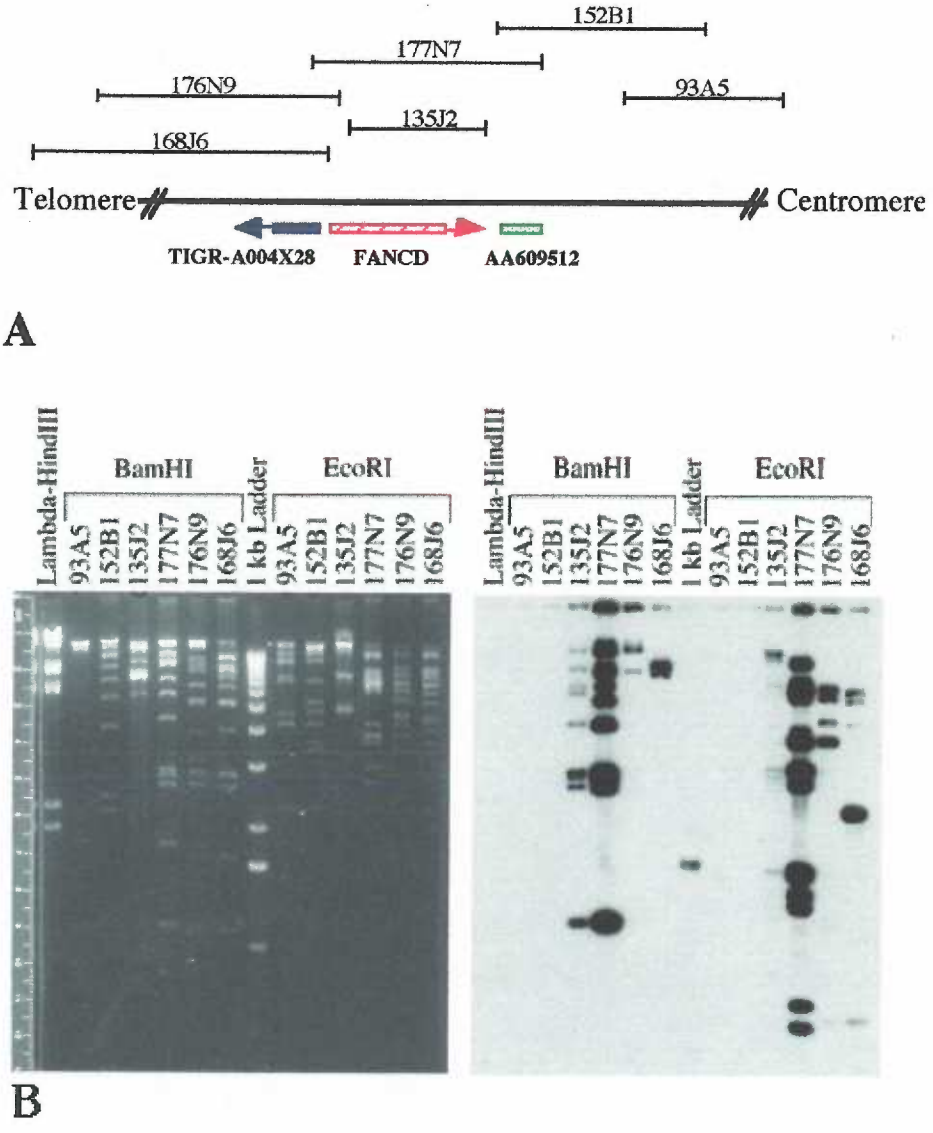


Figure 16. Transcript Map and Genomic Southern. **A)** Diagram of the BAC contig in the region of *FANCD*. The 5' end of *FANCD* is telomeric of the 3' end and TIGR-A004X28 is opposite. The transcriptional direction of AA609512 is unknown. **B)** The BACs from the region were digested with BamHI and EcoRI and probed with *FANCD*. BACs 135J2, 177N7, 176N9 and 168J6 all hybridize. The gene appears to be contained within 177N7.

Table 6. The Intron/Exon Junctions of *FANCD*

Exon	Size	5'-Donor site	Score	Intron	3'-Acceptor sit	Score	Exon
1	35	TCG gtgagtaagtg	87		gtttcccgattttgctctag GAA	85	2
2	98	CCA gtaagtatcta	83		gaaaatttttctattttcag AAA	83	3
3	142	TAG gtaatatttta	78		ctcttctttttttctgcatag CTG	88	4
4	69	AAA gtatgtatttt	81	159	atttttttaaatctccttaag ATA	78	5
5	105	CAG gtgtggagagg	86	375	gatttctttttttttttacag TAT	91	6
6	62	CAG gtaagactgtc	89		ccctatgtcttctttttttag CCT	86	7
7	54	AAA gtaagtggcgt	87		ttctcttcctaacattttag CAA	80	8
8	80	AAG gtaggcttatg	83	364	aatagtgtcttctactgcag GAC	85	9
9	126	CAG gtggataaacc	80		tctttttctaccattcacag TGA	86	10
10	89	AAG gtagaaaagac	76		tctgtgcttttaatttttag GTT	85	11
11	106	GAG gtatgctctta	80	387	ctaatatttactttctgcag GTA	87	12
12	102	AAG gtaaagagctc	85	342	tgctacttgtagttcctcag GAA	84	13
13	102	AAG gtgagatcttt	89	237	actctctcctgttttttcag GCA	92	14
14	37	AAG gtaatgttcat	82		tgcatatttattgacaatag GTG	73	15
15	145	TTA gtaagtgtcag	80		tctactcttccccactcaag GTT	86	16
16	136	CAG gtatgttgaaa	85		gttgactctcccctgtatag GAA	84	17
17	133	AAG gtatcttattg	77		tggcatcatttttttccacag GGC	89	18
18	112	CAG gttagaggcaa	83		tcttcatcatctcattgcag GAT	87	19
19	111	CAG gtacacgtgga	82		aaaaaattctttgtttttag AAG	79	20
20	62	CAG gtgagttcttt	93		attcttcctctttgctccag GTG	93	21
21	121	CTG gtaaagccaat	81	445	tgtttgtttgcttcctgaag GAA	85	22
22	75	AGG gtaggtattgt	84	300	attctggtttttctccgcag TGA	88	23
23	148	AAA gtcagtatagt	73		aatttatttctccttctcag ATT	89	24
24	102	TAG gtatgggatga	84	370	aaatgtttgttctctctcag ATT	86	25
25	117	GAG gtgagcagagt	88		atgtaatttgtactttgcag ATT	82	26
26	110	CAG gtaagagaagt	89		cagcctgctgtttgtttcag TCA	81	27
27	112	TAG gtaagtatgtt	90	272	ttctctttttaatataaaag AAA	73	28
28	111	AAG gtattggaatg	78		ttgctgtgacttccccatag GAG	85	29
29	145	GAA gtaagtgacag	85		tcctttcctccatgtgacag GCT	84	30
30	118	AAG gttagtgtagg	86		taactctgcatttattatag AAC	80	31
31	130	CAG gtcagaagcct	82	118	aaaatcatttttatttttag TGT	79	32
32	120	TTG gtaagtatgtg	85		tcttaccttgacttccttag GAG	85	33
33	112	CAG gtgagtcataa	90		tttttcttgtctccttacag CCA	91	34
34	132	TTG gtgatgggcct	73		tttgtcttcttttctaacag CTT	89	35
35	95	CTG gtgagatgttt	84	286	atatttgactctcaatgcag TAT	78	36
36	124	CAG gtaagggagtt	92		atgcttttcccgtcttctag GCA	88	37
37	95	CAG gtgagtaagat	92		catatatttggctgccccag ATT	81	38
38	73	AAG gtgagtatgga	93		cttgtctttcacctctccag GTA	93	39
39	40	AAG gtgagagattt	89		agtgtgtctctcttcttcag TAT	86	40
40	76	CGG gtaagagctaa	86		tataaacttattggttatag GAA	77	41
41	76	AAG gtaagaagggg	91		tgttatttatttccattcag ATT	86	42
42	148	CAG gtaagccttgg	91		cttggtccattcacatttag GGT	80	43
43	229	CCA taa + 3'UTR					

Table 7: PCR Primers to Amplify the 43 Exons of *FANCD*

Exon	Primer Name	Primer Sequence(5'->3')	Product Size (bp)
1	MG914	F: CTAGCACAGAACTCTGCTGC	372
	MG837	R: CTAGCACAGAACTCTGCTGC	
2	MG746	F: CTTCAGCAACAGCGAAGTAGTCTG	422
	MG747	R: GATTCTCAGCACTTGAAAAGCAGG	
3	MG773	F: GGACACATCAGTTTTCCTCTC	309
	MG789	R: GAAAACCCATGATTCAGTCC	
4-5	MG816	F: TCATCAGGCAAGAAACTTGG	467
	MG803	R: GAAGTTGGCAAAACAGACTG	
6	MG804	F: GAGCCATCTGCTCATTTCTG	283
	MG812	R: CCCGCTATTTAGACTTGAGC	
7	MG775	F: CAAAGTGTTTATTCCAGGAGC	343
	MG802	R: CATCAGGGTACTTTGAACATTC	
8-9	MG727	F: TTGACCAGAAAGGCTCAGTTCC	640
	MG915	R: AGATGATGCCAGAGGGTTTATCC	
10	MG790	F: TGCCCAGCTCTGTTCAAACC	222
	MG774	R: AGGCAATGACTGACTGACAC	
11	MG805	F: TGCCCGTCTATTTTTGATGAAGC	392
	MG791	R: TCTCAGTTAGTCTGGGGACAG	
12-14	MG751	F: TCATGGTAGAGAGACTGGACTGTGC	988
	MG752	R: GGAGGGGGGAGAAAGAAAGAC	
15	MG777	F: AGCTGCTGTTTCATTGTAGC	318
	MG793	R: CAATGGGTTTGAACAATGGACTG	
16	MG755	F: AACGTGGAATCCCATTGATGC	379
	MG730	R: TTTCTGTGTTCCCTCCTTGC	
17	MG794	F: GATGGTCAAGTTACACTGGC	382
	MG778	R: CACCTCCCACCAATTATAGTATTC	
18	MG808	F: CTATGTGTGTCTCTTTTACAGGG	234
	MG817	R: AATCTTTCCCACCATATTGC	
19	MG779	F: CATACCTTCTTTTGCTGTGC	199
	MG795	R: CCACAGAAGTCAGAATCTCCACG	
20	MG731	F: TGTAACAAACCTGCACGTTG	632
	MG732	R: TGCTACCCAAGCCAGTAGTTTCC	
21	MG788	F: GAGTTTGGGAAAGATTGGCAGC	232
	MG772	R: TGTAGTAAAGCAGCTCTCATGC	
22-23	MG733	F: CAAGTACACTCTGCACTGCC	652
	MG758	R: TGACTCAACTTCCCCACCAAGAG	
24-25	MG736	F: CTCCCTATGTACGTGGAGTAATAC	732
	MG737	R: GGGAGTCTTGTGGGAACTAAG	
26	MG780	F: TTCATAGACATCTCTCAGCTCTG	284
	MG759	R: GTTTTGGTATCAGGGAAAGC	
27-28	MG760	F: AGCCATGCTTGGAATTTTGG	653
	MG781	R: CTCACTGGGATGTCACAAAC	
29	MG740	F: GGTCTTGATGTGTGACTTGTATCCC	447
	MG741	R: CCTCAGTGTCACAGTGTTCTTTGTG	
30	MG809	F: CATGAAATGACTAGGACATTCC	281

	MG797	R: CTACCCAGTGACCCAAACAC	
31-32	MG761	F: CGAACCCTTAGTTTCTGAGACGC	503
	MG742	R: TCAGTGCCTTGGTGACTGTC	
33	MG916	F: TTGATGGTACAGACTGGAGGC	274
	MG810	R: AAGAAAGTTGCCAATCCTGTTCC	
34	MG762	F: AGCACCTGAAAATAAGGAGG	343
	MG743	R: GCCCAAAGTTTGTAAGTGTGAG	
35-36	MG787	F: AGCAAGAATGAGGTCAAGTTC	590
	MG806	R: GGGAAAAACTGGAGGAAAGAACTC	
37	MG818	F: AGAGGTAGGGAAGGAAGCTAC	233
	MG813	R: CCAAAGTCCACTTCTTGAAG	
38	MG834	F: GATGCACTGGTTGCTACATC	275
	MG836	R: CCAGGACACTTGGTTTCTGC	
39	MG839	F: ACACTCCCAGTTGGAATCAG	370
	MG871	R: CTTGTGGGCAAGAAATTGAG	
40	MG829	F: TGGGCTGGATGAGACTATTC	223
	MG870	R: CCAAGGACATATCTTCTGAGCAAC	
41	MG820	F: TGATTATCAGCATAGGCTGG	271
	MG811	R: GATCCCCCAATAGCAACTGC	
42	MG763	F: CATTCAGATTCACCAGGACAC	227
	MG782	R: CCTTACATGCCATCTGATGC	
43	MG764	F: AACCTTCTCCCCTATTACCC	435
3'UTR	MG835	R: GGAAAATGAGAGGCTATAATGC	

Characterization of the 5' region. Approximately 1100 base pairs around exon 1 were sequenced from BAC 177N7. Analysis of this region failed to locate any TATA and CAAT boxes, but did identify two CpG islands, one 413 bp upstream of the predicted transcription start site and one in the first intron. Interestingly, it was found that the first exon of TIGR-A004X28 is only 97 bp upstream of exon one of *FANCD*. Several possible binding sites for transcription factors were identified using TRANSFAC (Version 3.3) in the region between *FANCD* and TIGR-A004X28, however, it is impossible at this time to determine which gene they are regulating.

Identification of a murine homolog of *FANCD*. As a first step in the production of a knock-out mouse for *FANCD*, I isolated a mouse homolog of the human cDNA. The

human cDNA sequence was used to search the mouse EST database. Several high scoring clones were identified. Primers were designed from these EST sequences and used for 5' RACE on the Marathon-Ready mouse testis cDNA library (Clontech). The full-length sequence was obtained. The mouse *FANCD* gene has an open reading frame of 4,350 bp and encodes a 1,452 amino acid protein. The mouse and human genes are 79% conserved at the DNA level and their proteins are 73% identical. Both the serine (S126) and arginine (R1236) which are mutated in PD20 are conserved between the two, where as the L/P714 polymorphism is not.

It is worth mentioning here, that a possible homolog in *Arabidopsis thaliana* has also been identified. BLAST searches revealed a homology with a P-value of 10^{-46}, to a hypothetical protein in the plant. The sequence in the database is the result of a genome sequences project and therefore, no function has been identified for the protein. The protein alignments appear to be in clusters towards the amino-terminus, but due to sequencing errors, the homology may be lost on the carboxy-terminus (Figure 17). Homologs in other organisms such as yeast and Drosophilia have not been identified.

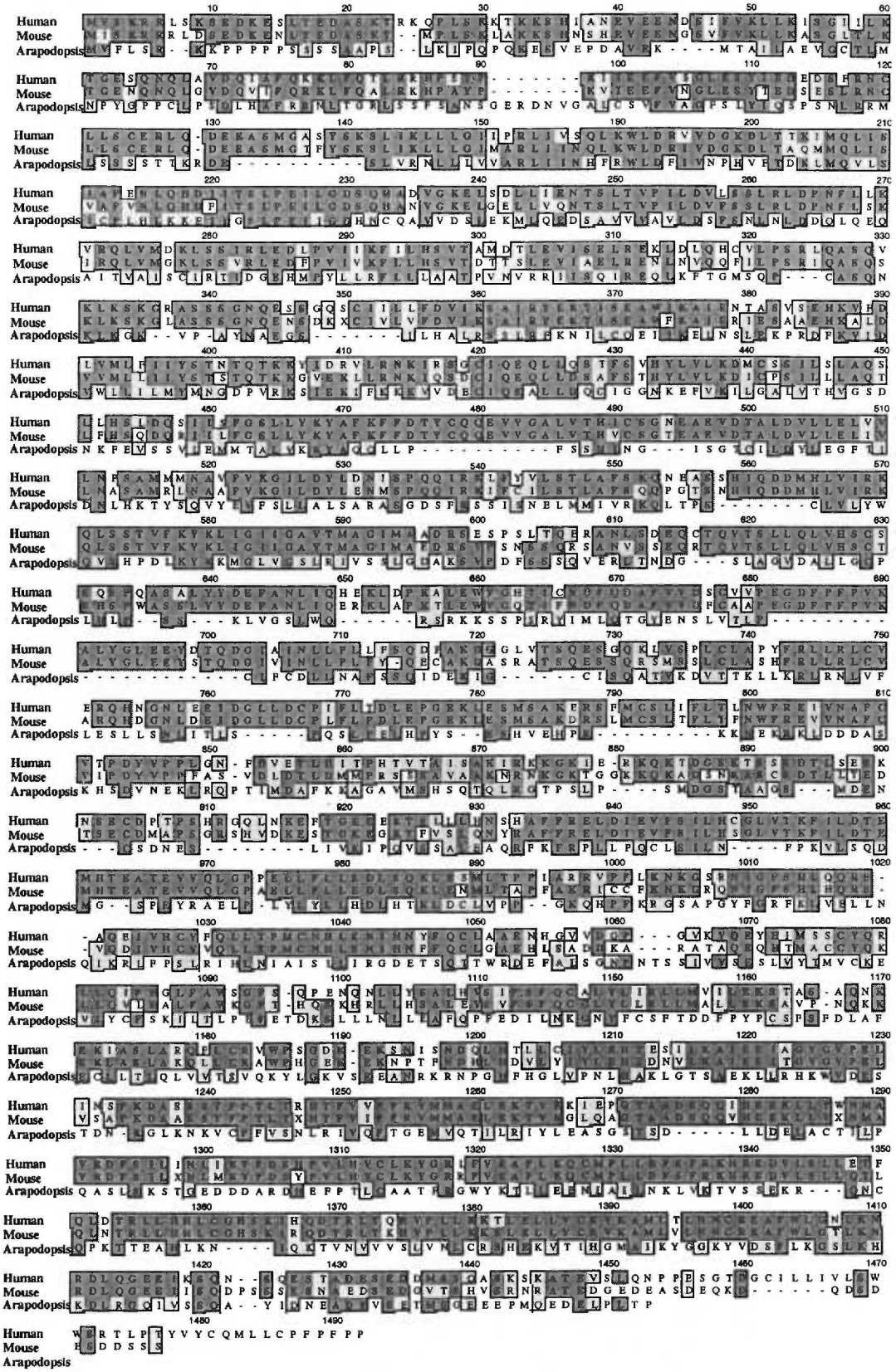


Figure 17: Protein alignment of the human, mouse and arapodopsis homologs.

Chapter Four:
Discussion and Conclusions

To date, over 600 human disease genes have been identified. The cloning of a human disease gene is often the first step to a better understanding of the molecular basis of the disease. In turn, a better understanding at the molecular level can potentially improve treatment strategies. This thesis describes the positional cloning and initial characterization of the Fanconi anemia complementation group D gene (*FANCD*). The isolation of all the genes involved in Fanconi anemia (FA) is a high priority in the field not only because they may provide clues to a possible molecular mechanism but also because they may offer a cure through gene therapy.

Sub-localization of the *FANCD* gene

Currently, there are at least 8 different genes involved in FA (Joenje, et al., 1997). Three of these genes have been isolated using an expression cloning method and one out of the three (FANCA) was also positionally cloned (Buchwald, et al., 1992, consortium, 1996, Lo Ten Foe, et al., 1996a, de Winter, et al., 1998). It is worth mentioning that FANCA is the most prevalent complementation group in people of Northern European descent, accounting for over 65% of FA patients (Duckworth-Rysiecki, et al., 1985, Strathdee, et al., 1992a, Jakobs, et al., 1996). Linkage analysis was, therefore, possible because of the availability of a large number of families. Unfortunately, the other groups typically include only a small number of families, making mapping of the gene by linkage analysis impossible. Expression cloning has been successful because of the high sensitivity of some FA cells to cross-linking agents, which allowed good selection of the complemented clones. Several drawbacks to this

method, however, exist. First, there are a large number of false positives due to the over-expression of growth enhancing genes. Second, there is a possibility that a full-length clone of the disease gene is not present in the expression library. And finally, over-expression of the disease gene could lead to cellular toxicity.

For two reasons we believed that neither the linkage analysis nor expression cloning approach would work well for cloning the *FANCD* gene. First *FANCD*$^-$ cells are not as sensitive to cross-linking as other groups, which makes selection for complemented clones difficult. Second, there are only five small families in the world, making linkage analysis impossible. For these reasons, we chose to use microcell-mediated chromosome transfer (MMCT) to map the *FANCD* gene. This method involves moving individual human chromosomes into cells so that all the genes introduced are single copy and regulated by their own promoters.

The first section of this thesis describes the sub-localization of the *FANCD* gene to a 200 kb region on chromosome 3p25.3 through the deletion analysis of several hybrids produced by MMCT. The section also describes the construction of a BAC contig that spans approximately 1-2 Mb in this region.

One of the most valuable tools we used to sub-localize the *FANCD* gene was the STS and EST databases developed by the human genome project. The STS database contains mapping information and primer sequence for a continuously growing number of polymorphic markers. The EST database is composed of bins of cDNAs (like baskets of apples) delineated by a specific set of markers that have been mapped to the human genome by both radiation hybrids and physical maps. By simply entering two marker

numbers, researchers can obtain a set of ESTs that have been localized to their region of interest. The database also contains sequence and homology information for each. We used these databases at first, to help us define the deletion in our original non-complemented clone and later to identify ESTs that mapped to the sub-localized region. These ESTs were vital for the construction of the BAC contig, as well as helping us identify candidate genes. The availability of these databases has clearly made the sub-localization of the gene possible.

Although there are several different genomic libraries available for building contigs, we chose to use BACs because of the advantages they offer. First, the size of the BACs (50-200 kb) easily allows the coverage of a 1-2 Mb region of genomic DNA as compared to cosmid libraries, yet they are still easily purified by a modified alkaline lysis preparation. Second, BACs are rarely chimeric as compared to YACs. In our study, we found only one BAC that was chimeric. BACs can also potentially be directly used for functional complementation of cell lines. Finally, the use of BACs as probes for FISH enabled us to determine their order in the contig as well as define the deletions in the additional non-complemented hybrids.

The CIT BAC library that was used to build the 3p25.3 contig, represents approximately a six-fold coverage of the human genome. It is interesting to note that in the middle of the contig the coverage dropped to less than one-fold. It was also noted that in the YAC/cosmid contig covering the *VHL* region, YACs in this same region were chimeric (Kuzmin, et al., 1994). A possible explanation for this observation might be that the region is inherently unstable, making it difficult to either clone the DNA of this

region or maintain it in a vector. Aside from deletion breakpoints associated with VHL, this same region has also been associated with translocation breakpoints (Schmidt, et al., 1998). These observations are in agreement with the region being unstable.

During the construction of the contig, several previously cloned genes were mapped to the region, including the oxytocin receptor, calmodulin kinase-1, the interferon receptor-associated kinase-2 (*IRAK-2*), the human plasma membrane calcium ATPase isoform 2 gene (*ATP2B2*) and the human homolog to the 8-oxoguanine glycosylase gene. *hOGG1* was considered as a primary candidate gene because of its potential role in DNA repair of oxidative damage. Since FA cells have been shown to be sensitive to oxygen, the involvement of this type of protein in the disease is easily justified. After sequence and functional analysis the gene was, however, ruled out as a candidate.

One could also postulate how *IRAK-2* could potentially be an FA gene. Since both human and mouse FA cells are sensitive to interferon-gamma, it has been hypothesized that the FA proteins are involved in the signaling pathway of this cytokine (Rathbun, et al., 1997). It is easy, therefore, to imagine the role a receptor-associated kinase might play in this type of pathway. Our functional data, however, demonstrated that this gene is not *FANCD*.

Overall, this study demonstrated the power behind the use of somatic cell genetics to localize genes to a very defined region on a chromosome. It also showed how the advances in the human genome project has made the identification of candidate genes remarkably easier. And finally, it describes a methodical approach for constructing a

genomic contig using BACs and subsequently using them as tools for FISH analysis. Clearly, this BAC contig will also facilitate the cloning of many other genes that map to the region.

Analysis of candidate cDNAs

The second section of this thesis describes the analysis of three EST as candidate genes for *FANCD*. The first of these genes, AA609512, was easily eliminated based on its specific expression in the testis. Since the three other cloned FA genes are ubiquitously expressed and the disease affects several organs including the hematopoietic system, we assumed that *FANCD* would not be tissue specific. The expression pattern of AA609512, therefore, did not fit the profile.

At the time we began our analysis of the three candidate genes, BLAST searches with TIGR-A004X28 revealed a single homology to a fat-specific protein. Based on this finding, we chose not to investigate it further. A more recent BLAST search has, however, revealed an additional homology to a cell death activator protein (CIDE). CIDEs are a novel family of proteins that have been shown to induce apoptosis. Interestingly, genomic sequence analysis has shown that the first exon of TIGR-A004X28 is located just 70 bp upstream of the transcriptional start site of *FANCD*. Because of the proximity of the two genes, it is possible that their regulatory regions overlap. Hence, the regulation of the two genes may be affected by promotor competition. This observation is potentially important given that the over-expression of FANCD kills cells, possibly by inducing apoptosis.

The third gene, SGC34603, was shown to have a low ubiquitous expression pattern, matching that of the other FA genes. Based on these results, it became our best candidate for being the *FANCD* gene. Sequence analysis of the gene from the PD20 patient DNA revealed two sequence alterations. Although neither mutation was obviously deleterious, neither change was present in 400 control chromosomes. Polymorphisms are defined as sequence variations with a population frequency of at least 1%. Therefore, neither alteration can be considered a polymorphism.

Mutation analysis of *FANCD*

The maternal mutant allele carried a single base pair substitution at position 376 in exon 5. Although this is a missense mutation, we believe that its deleterious effect is due to the disruption of the splice donor site of exon 5. Every cDNA clone that was sequenced with this mutation had an additional thirteen bases of intronic sequence downstream of exon 5. The most likely explanation for this is that the normal splice site cannot be recognized and the splicing machinery uses a cryptic intronic site instead. The addition of thirteen bases disrupts the open reading frame and would result in a severely truncated protein.

The second mutation was also a single base pair change causing a missense mutation (R1236H). Interestingly though, all of the cDNA clones from this allele had exon 37 deleted. It could be that the mutation affects the splicing of the exon even though it is in the middle. It was also observed that there appeared to be a skewed ratio of maternal to paternal alleles (more maternal cDNAs) when the cDNA clones were

analyzed. Therefore, it is also possible that the mutation affects the stability of the mRNA and only cDNA clones that have the mutation spliced-out, can be observed. Cases of both of these situations have been previously reported (Mohsen, et al., 1998, Vockley, et al., 1998). Unfortunately, both of these mutations appear to disrupt the proper translation of the protein and will probably not lead us to any clues about functional domains. It is possible that the identification of mutations in the four other *FANCD* $^{-}$patients will give us better clues in this respect.

Somatic reversion studies in the PD20 cell line

We recently received a bone marrow sample from the PD20 patient. The patient had previously suffered from pancytopenia. However, at his most recent check-up his blood counts had begun to rise. His physician sent us a new sample to analyze for chromosome breakage and determine whether a somatic reversion in the patients hematopoietic system had occurred. The patient's bone marrow sample remained sensitive to DNA cross-linking agents, however EBV-transformed lymphoblasts from the sample were less sensitive. This result was not surprising since it has been documented that the lymphoblasts of approximately 25% of FA patients are mosaic as demonstrated by the presence of both phenotypically corrected and non-corrected cells (Lo Ten Foe, et al., 1997). Theoretically, mosaicism could originate from genetic changes at the disease locus or at modifier genes capable of alleviating the disease phenotype. Since the latter has never been demonstrated in FA, the mosaicism is most likely caused by a somatic reversion at the disease locus. Somatic reversion due to mitotic intragenic recombination

has been described in Bloom syndrome, ataxia telangiectasia and FA (Cohen, et al., 1979, Groden, et al., 1990, Ellis, et al., 1995, Lo Ten Foe, et al., 1997). This is the most common cause of mosaicism, especially in compound heterozygotes, but is not the only mechanism. Other mechanisms include point mutation reversions and genomic insertions and deletions.

The results of our reversion study show that the mosaicism in the PD20 lymphoblasts is most likely due to an intragenic mitotic recombination event. Two types of recombination could be inferred from the genotypes of the reverted cells. A single intragenic cross-over between the maternally and paternally inherited mutations could have occurred. If this were the case, the reversion would be associated with haplotype switching for markers located distally to the disease gene. The reversion could also have been caused by a gene conversion which would result in segregants losing one pathogenic mutation. Without further analysis, however, it is impossible to determine which mode of recombination occurred in the PD20 lymphoblasts.

The fact that such a high percentage of FA patients have mosaic lymphocytes suggests that an underlying force may be driving the phenotypic reversion. If FA cells are defective in double-strand break repair, it is possible that they may have to rely on homologous recombination to repair them. If this were the case, the cells would have an increased rate of homologous recombination which would increase the chances of a somatic reversion. Indeed, this has been reported (Thyagarajan and Campbell, 1997). It is also worth noting that if the reversion occurred in a stem cell, the resulting hematopietic progenitor cells would have a proliferative advantage over poorly growing

non-corrected progenitors. This could eventually lead to the repopulation of the bone marrow by the corrected cells.

Cytotoxicity of FANCD

Our results have shown conclusively that the *FANCD* gene is toxic to cells when overexpressed. There are many mechanisms that could explain this phenotype. The most obvious of which is that FANCD induces apoptosis as a cell cycle regulator. Since FA cells have been shown to be sensitive to interferon-gamma, it is possible that FANCD plays a role in the apoptotic signaling pathway of this cytokine. The gene may also play a more direct role such as a regulator of DNA fragmentation. Alternatively, FANCD could directly bind DNA such as an endonuclease in a repair complex. When overexpressed an endonuclease would definitely destroy a cell. A study with FA cells reported the possible involvement of a DNA damage recognition complex that binds DNA containing interstrand cross-links (Hang, et al., 1993). It is possible that FANCD is part of this complex and acts as an endonuclease involved in the repair of double-strand breaks. Given that there is evidence to support the involvement of a FA protein in either one of these functions, both areas will need to be investigated.

It is important to note here that because the *FANCD* gene is toxic when overexpressed it would not have been isolated by an expression cloning technique. For this reason, several approaches should always be used to identify disease genes. Unfortunately, the cytotoxicity phenotype also means that gene therapy for group D patients will be difficult.

Genomic structure of *FANCD*

The final section of this thesis describes the genomic structure of *FANCD*. The gene has an open reading frame of 4,413 bp comprised of 43 exons. The start methionine is in the second exon, therefore, the first exon contributes to the 68 bp of 5' UTR. Although we observed several splice variants, we found no evidence that the abberent splicing was due to poor splice recognition sites. The consensus sequence for all the splice donor and acceptor sites was appropriate. Analysis of the region upstream of the transcriptional start site did not reveal a TATA box or any other obvious promotor. As mention previously, the transcription start site of another gene, TIGR-A004X28, is located just upstream of *FANCD*. The implications of this have been previously discussed. The availability of the genomic sequence will now allow us to do mutation analysis on the remaining *FANCD* $^-$ patients from which we only have genomic DNA.

The deduced amino acid sequence of FANCD revealed a putative transmembrane domain. In addition, the protein appears to be extremely hydrophobic.

Identification of *FANCD* homologs

BLAST searches using the *FANCD* sequence has revealed a possible homolog in both mouse and *Arabidopsis thaliana*. We used the mouse EST sequence to clone the full-length gene. The cloning of the murine homolog will facilitate the production of a mouse knock-out and will provide a necessary tool for *in situ* hybridizations.

The identification of an arabidopsis homolog is significant because it is the first FA gene identified that has a homolog in a non-vertebrate species. Although there does

not appear to be a homolog in yeast or *E.coli*, it does not exclude the possibility that one exists in other model organisms, such as *Drosophila* or *Aspergillus nidulans*. Identification of a homolog in one of these organisms would greatly facilitate FA research.

Future Studies

The cloning of the *FANCD* gene opens the door for a plethora of future experiments. These include the generation of a mouse knock-out which can be used for phenotypic studies. The production of combo-mice for the different FA groups may also provide clues to the epistatic relationships of the genes. It will also be important to determine how the over-expression of FANCD kills the cell. The results of such experiments will most likely provide a clue to what the endogenous function of the protein is. Finally, the generation of an antibody could facilitate cellular localization studies as well as co-immunoprecipitation assays to identify interacting proteins.

References

Aburatani, H., Hippo, Y., Ishida, T., Takashima, R., Matsuba, C., Kodama, T., Takao, M., Yasui, A., Yamamoto, K. and Asano, M. (1997). Cloning and characterization of mammalian 8-hydroxyguanine-specific DNA glycosylase/apurinic, apyrimidinic lyase, a functional mutM homologue. Cancer Res *57,* 2151-6.

Alter, B. P. (1993). Fanconi's anaemia and its variability. Br J Haematol *85,* 9-14.

Alter, B. P. (1995). Hematologic abnormalities in Fanconi anemia. Blood *85,* 1148-9.

Alter, B. P. (1996). Fanconi's anemia and malignancies. Am J Hematol *53,* 99-110.

Alter, B. P., Knobloch, M. E. and Weinberg, R. S. (1991). Erythropoiesis in Fanconi's anemia. Blood *78,* 602-8.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. J Mol Biol *215,* 403-10.

Auerbach, A. D. (1993). Fanconi anemia diagnosis and the diepoxybutane (DEB) test. Exp Hematol *21,* 731-3.

Auerbach, A. D., Rogatko, A. and Schroeder-Kurth, T. M. (1989). International Fanconi Anemia Registry: relation of clinical symptoms to diepoxybutane sensitivity. Blood *73,* 391-6.

Auerbach, A. D. and Wolman, S. R. (1976). Susceptibility of Fanconi's anaemia fibroblasts to chromosome damage by carcinogens. Nature *261,* 494-6.

Bagnara, G. P., Strippoli, P., Bonsi, L., Brizzi, M. F., Avanzi, G. C., Timeus, F., Ramenghi, U., Piaggio, G., Tong, J., Podesta, M. and et, a. (1992). Effect of stem cell

factor on colony growth from acquired and constitutional (Fanconi) aplastic anemia. Blood *80,* 382-7.

Berger, R., Le-Coniat, M. and Gendron, M. C. (1993). Fanconi anemia. Chromosome breakage and cell cycle studies. Cancer Genet Cytogenet *69,* 13-6.

Bigelow, S. B., Rary, J. M. and Bender, M. A. (1979). G2 chromosomal radiosensitivity in Fanconi's anemia. Mutat Res *63,* 189-99.

Broxmeyer, H. E., Douglas, G. W., Hangoc, G., Cooper, S., Bard, J., English, D., Arny, M., Thomas, L. and Boyse, E. A. (1989). Human umbilical cord blood as a potential source of transplantable hematopoietic stem/progenitor cells. Proc Natl Acad Sci U S A *86,* 3828-32.

Broxmeyer, H. E., Gluckman, E., Auerbach, A., Douglas, G. W., Friedman, H., Cooper, S., Hangoc, G., Kurtzberg, J., Bard, J. and Boyse, E. A. (1990). Human umbilical cord blood: a clinically useful source of transplantable hematopoietic stem/progenitor cells. Int J Cell Cloning *1,* 76-89.

Buchwald, M. and Moustacchi, E. (1998). Is Fanconi anemia caused by a defect in the processing of DNA damage? Mutat Res *408,* 75-90.

Bunz, F., Dutriaux, A., Lengauer, C., Waldman, T., Zhou, S., Brown, J. P., Sedivy, J. M., Kinzler, K. W. and Vogelstein, B. (1998). Requirement for p53 and p21 to sustain G2 arrest after DNA damage. Science *282,* 1497-501.

Butturini, A., Gale, R. P., Verlander, P. C., Adler-Brecher-B, G. A. P. and Auerbach, A. D. (1994). Hematologic abnormalities in Fanconi anemia: an InternationalFanconi Anemia Registry study. Blood *84,* 1650-5.

Chaganti, R. S. and Houldsworth, J. (1991). Fanconi anemia: a pleotropic mutation with multiple cellular and developmental abnormalities. Ann Genet *34,* 206-11.

Chen, M., Tomkins, D. J., Auerbach, W., McKerlie, C., Youssoufian, H., Liu, L., Gan, O., Carreau, M., Auerbach, A., Groves, T., Guidos, C. J., Freedman, M. H., Cross, J., Percy, D. H., Dick, J. E., Joyner, A. L. and Buchwald, M. (1996). Inactivation of Fac in mice produces inducible chromosomal instability and reduced fertility reminiscent of Fanconi anaemia. Nat Genet *12,* 448-51.

Ching-Ying-Wong, J., Alon, N. and Buchwald, M. (1997). Cloning of the bovine and rat Fanconi anemia group C cDNA. Mamm Genome *8,* 522-5.

Clarke, A. A., Philpott, N. J., Gordon-Smith, E. C. and Rutherford, T. R. (1997). The sensitivity of Fanconi anaemia group C cells to apoptosis induced by mitomycin C is due to oxygen radical generation, not DNA crosslinking. Br J Haematol *96,* 240-7.

Cohen, M. M., Fruchtman, C. E., Simpson, S. J. and Martin, A. O. (1982). The cytogenetic response of Fanconi's anemia lymphoblastoid cell lines to various clastogens. Cytogenet Cell Genet *34,* 230-40.

Cohen, M. M., Sagi, M., Ben-Zur, Z., Schaap, T., Voss, R., Kohn, G. and Ben-Bassat, H. (1979). Ataxia teleangiectasie: chromosomal stability in continous lymphoblastoid cell lines. Cytogenet Cell Genet *23,* 44-52.

Collod, G., Babron, M. C., Jondeau, G., Coulon, M., Weissenbach, J., Dubourg, O., Bourdarias, J. P., Bonaiti-Pellie, C., Junien, C. and Boileau, C. (1994). A second locus for Marfan syndrome maps to chromosome 3p24.2-p25 [see comments]. Nat Genet *8,* 264-8.

consortium, F. a. (1996). Positional cloning of the Fanconi anaemia group A gene. Nat. Genet. *14,* 324-28.

Cowling, G. J. and Dexter, T. M. (1994). Apoptosis in the haemopoietic system. Philos Trans R Soc Lond B Biol Sci *345,* 257-63.

Cumming, R. C., Liu, J. M., Youssoufian, H. and Buchwald, M. (1996). Suppression of apoptosis in hematopoietic factor-dependent progenitor cell lines by expression of the FAC gene. Blood *88,* 4558-67.

D'Andrea, A. D. and Grompe, M. (1997). Molecular biology of Fanconi anemia: implications for diagnosis and therapy. Blood *90,* 1725-36.

Daneshbod-Skibba, G., Martin, J. and Shahidi, N. T. (1980). Myeloid and erythroid colony growth in non-anaemic patients with Fanconi's anaemia. Br J Haematol *44,* 33-8.

Davies, S. M., Khan, S., Wagner, J. E., Arthur, D. C., Auerbach, A. D., Ramsay, N. K. and Weisdorf, D. J. (1996). Unrelated donor bone marrow transplantation for Fanconi anemia. Bone Marrow Transplant *17,* 43-7.

de Winter, J. P., Waisfisz, Q., Rooimans, M. A., van Berkel, C. G., Bosnoyan-Collins, L., Alon, N., Carreau, M., Bender, O., Demuth, I., Schindler, D., Pronk, J. C., Arwert, F., Hoehn, H., Digweed, M., Buchwald, M. and Joenje, H. (1998). The Fanconi anaemia group G gene FANCG is identical with XRCC9. Nat Genet *20,* 281-3.

Digweed, M., Zakrzewski-Ludcke, S. and Sperling, K. (1988). Fanconi's anaemia: correlation of genetic complementation group with psoralen/UVA response. Hum Genet *78,* 51-4.

Doyle, A., Griffiths, J. B. and Newell, D. G. (1993). Cell & tissue culture : laboratory procedures, (Chichester [England] ; New York: Wiley)

Drumheller, T., McGillivray, B. C., Behrner, D., MacLeod, P., McFadden, D. E., Roberson, J., Venditti, C., Chorney, K., Chorney, M. and Smith, D. I. (1996). Precise localisation of 3p25 breakpoints in four patients with the 3p- syndrome. J Med Genet *33,* 842-7.

Duckworth-Rysiecki, G., Cornish, K., Clarke, C. A. and Buchwald, M. (1985). Identification of two complementation groups in Fanconi anemia. Somat Cell Mol Genet *11,* 35-41.

Duckworth-Rysiecki, G. and Taylor, A. M. R. (1985). Effects of ionizing radiation on cells from Fanconi's anemia pateints. Cancer Res. *45,* 2015-2020.

Ellis, N. A., Lennon, D. J., Proytcheva, M., Alhadeff, B., Henderson, E. E. and German, J. (1995). Somatic intragenic recombination within the mutated locus BLM can correct the high sister-chromatid exchange phenotype of Bloom syndrome cells. Am J Hum Genet *57,* 1019-27.

Escarceller, M., Buchwald, M., Singleton, B. K., Jeggo, P. A., Jackson, S. P., Moustacchi, E. and Papadopoulo, D. (1998). Fanconi anemia C gene product plays a role in the fidelity of blunt DNA end-joining. Journal of Molecular Biology *279,* 375-85.

Evans, D. G., Rees, H. C., Spreadborough, A., Campbell, D. J., Gau, G. S., Pickering, E., Hamilton, S. and Clayton-Smith, J. (1994). Radial ray defects, renal ectopia, duodenal atresia and hydrocephalus: the extended spectrum for Fanconi anaemia. Clin Dysmorphol *3,* 200-6.

Fanconi, G. (1967). Familial constitutional panmyelocytopathy, Fanconi's anemia (F.A.) I. Clinical aspects. Semin. Hematol. *4,* 233-240.

Flowers, M. E., Zanis, J., Pasquini, R., Deeg, H. J., Ribeiro, R., Longton, G., Medeiros, C. R., Doney, K., Sanders, J., Bryant, J. and et, a. (1996). Marrow transplantation for Fanconi anaemia: conditioning with reduced doses of cyclophosphamide without radiation. Br J Haematol *92,* 699-706.

Fournier, R. E. (1981). A general high-efficiency procedure for production of microcell hybrids. Proc Natl Acad Sci U S A *78,* 6349-53.

Fujiwara, Y. and Tatsumi, M. (1977). Cross-link repair in human cells and its possible defect in Fanconi's anemia cells. J Mol Biol *113,* 635-49.

Gavish, H., dos-Santos, C. C. and Buchwald, M. (1993). A Leu554-to-Pro substitution completely abolishes the functional complementing activity of the Fanconi anemia (FACC) protein. Hum Mol Genet *2,* 123-6.

German, J. and Puglatta, C. (1966). Chromosomal breakage in diploid cell lines from Bloom's syndrome and Fanconi's anemia. Ann Genet *9,* 143.

Gibson, R. A., Buchwald, M., Roberts, R. G. and Mathew, C. G. (1993a). Characterization of the exon structure of the fanconi anemia group C gene by vectorette PCR. Hum Mol Genet *2,* 35-38.

Gibson, R. A., Hajianpour, A., Murer-Orlando, M., Buchwald, M. and Mathew, C. G. (1993b). A nonsense mutation and exon skipping in the Fanconi anaemia group C gene. Human Molecular Genetics *2,* 797-799.

Gibson, R. A., Morgan, N. V., Goldstein, L. H., Pearson, I. C., Kesterton, I. P., Foot, N. J., Jansen, S., Havenga, C., Pearson, T., de-Ravel, T. J., Cohn, R. J., Marques, I. M., Dokal, I., Roberts, I., Marsh, J., Ball, S., Milner, R. D., Llerena, J. C., Jr., Samochatova, E., Mohan, S. P., Vasudevan, P., Birjandi, F., Hajianpour, A., Murer-Orlando, M. and Mathew, C. G. (1996). Novel mutations and polymorphisms in the Fanconi anemia group C gene. Hum Mutat *8,* 140-8.

Gille, J. J., Wortelboer, H. M. and Joenje, H. (1987). Antioxidant status of Fanconi anemia fibroblasts. Hum Genet *77,* 28-31.

Glanz, A. and Fraser, F. C. (1982). Spectrum of anomalies in Fanconi anemia. J. Med. Genet. *19,* 412-416.

Gordon-Smith, E. C. and Rutherford, T. R. (1991). Fanconi anemia: constitutional aplastic anemia. Semin Hematol *28,* 104-12.

Groden, J., Nakamura, Y. and German, J. (1990). Molecular evidence that homologous recombination occurs in proliferating human somatic cells. Proc Natl Acad Sci U S A *87,* 4315-9.

Guillouf, C., Laquerbe, A., Moustacchi, E. and Papadopoulo, D. (1993). Mutagenic processing of psoralen monoadducts differ in normal and Fanconi anemia cells. Mutagenesis *8,* 355-61.

Guinan, E. C., Lopez, K. D., Huhn, R. D., Felser, J. M. and Nathan, D. G. (1994). Evaluation of granulocyte-macrophage colony-stimulating factor for treatment of pancytopenia in children with fanconi anemia. J Pediatr *124,* 144-50.

Hang, B., Yeung, A. T. and Lambert, M. W. (1993). A damage-recognition protein which binds to DNA containing interstrand cross-links is absent or defective in Fanconi anemia, complementation group A, cells. Nucleic Acids Res *21,* 4187-92.

Haribabu, B., Hook, S. S., Selbert, M. A., Goldstein, E. G., Tomhave, E. D., Edelman, A. M., Snyderman, R. and Means, A. R. (1995). Human calcium-calmodulin dependent protein kinase I: cDNA cloning, domain structure and activation by phosphorylation at threonine-177 by calcium-calmodulin dependent protein kinase I kinase. Embo J *14,* 3679-86.

Hartwell, L. H. and Weinert, T. A. (1989). Checkpoints: controls that ensure the order of cell cycle events. Science *246,* 629-34.

Heinrich, M. C., Hoatlin, M. E., Zigler, A. J., Silvey, K. V., Bakke, A. C., Keeble, W. W., Zhi, Y., Reifsteck, C. A., Grompe, M., Brown, M. G., Magenis, R. E., Olson, S. B. and Bagby, G. C. (1998). DNA cross-linker-induced G2/M arrest in group C Fanconi anemia lymphoblasts reflects normal checkpoint function. Blood *91,* 275-87.

Hirschhorn, R., Yang, D. R., Puck, J. M., Huie, M. L., Jiang, C. K. and Kurlandsky, L. E. (1996). Spontaneous in vivo reversion to normal of an inherited mutation in a patient with adenosine deaminase deficiency. Nat Genet *13,* 290-5.

Hoatlin, M. E., Christianson, T. A., Keeble, W. W., Hammond, A. T., Zhi, Y., Heinrich, M. C., Tower, P. A. and Bagby, G. C., Jr. (1998). The Fanconi anemia group C gene product is located in both the nucleus and cytoplasm of human cells. Blood *91,* 1418-25.

Hojo, E. T., van-Diemen, P. C., Darroudi, F. and Natarajan, A. T. (1995). Spontaneous chromosomal aberrations in Fanconi anaemia, ataxia telangiectasia fibroblast and Bloom's syndrome lymphoblastoid cell lines as detected by conventional cytogenetic analysis and fluorescence in situ hybridisation (FISH) technique. Mutat Res *334,* 59-69.

Hoshino, T., Wang, J., Devetten, M. P., Iwata, N., Kajigaya, S., Wise, R. J., Liu, J. M. and Youssoufian, H. (1998). Molecular chaperone GRP94 binds to the Fanconi anemia group C protein and regulates its intracellular expression. Blood *91,* 4379-86.

Ianzano, L., D'Apolito, M., Centra, M., Savino, M., Levran, O., Auerbach, A. D., Cleton-Jansen, A. M., Doggett, N. A., Pronk, J. C., Tipping, A. J., Gibson, R. A., Mathew, C. G., Whitmore, S. A., Apostolou, S., Callen, D. F., Zelante, L. and Savoia, A. (1997). The genomic organization of the Fanconi anemia group A (FAA) gene. Genomics *41,* 309-14.

Ioannou, P. A., Amemiya, C. T., Garnes, J., Kroisel, P. M., Shizuya, H., Chen, C., Batzer, M. A. and de Jong, P. J. (1994). A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. Nat Genet *6,* 84-9.

Jakobs, P. M., Fiddler-Odell, E., Reifsteck, C., Olson, S., Moses, R. E. and Grompe, M. (1997). Complementation group assignments in Fanconi anemia fibroblast cell lines from North America. Somat Cell Mol Genet *23,* 1-7.

Jakobs, P. M., Sahaayaruban, P., Saito, H., Reifsteck, C., Olson, S., Joenje, H., Moses, R. E. and Grompe, M. (1996). Immortalization of four new Fanconi anemia fibroblast cell lines by an improved procedure. Somat Cell Mol Genet *22,* 151-7.

Joenje, H. (1996). Fanconi anaemia complementation groups in Germany and The Netherlands. European Fanconi Anaemia Research group. Hum Genet *97,* 280-2.

Joenje, H., Frants, R. R., Arwert, F., de Bruin, G. J., Kostense, P. J., van de Kamp, J. J., de Koning, J. and Eriksson, A. W. (1979). Erythrocyte superoxide dismutase deficiency in Fanconi's anaemia established by two independent methods of assay. Scand J Clin Lab Invest *39,* 759-64.

Joenje, H., Lo-ten-Foe, J. R., Oostra, A. B., van-Berkel, C. G., Rooimans, M. A., Schroeder-Kurth, T., Wegner, R. D., Gille, J. J., Buchwald, M. and Arwert, F. (1995a). Classification of Fanconi anemia patients by complementation analysis: evidence for a fifth genetic subtype. Blood *86,* 2156-60.

Joenje, H., Mathew, C. and Gluckman, E. (1995b). Fanconi anaemia research: current status and prospects. Eur J Cancer *2,* 268-72.

Joenje, H., Oostra, A. B., Wijker, M., di Summa, F. M., van Berkel, C. G., Rooimans, M. A., Ebell, W., van Weel, M., Pronk, J. C., Buchwald, M. and Arwert, F. (1997). Evidence for at least eight Fanconi anemia genes. Am J Hum Genet *61,* 940-4.

Joenje, H., Youssoufian, H., Kruyt, F. A., dos-Santos, C. C., Wevrick, R. and Buchwald, M. (1995c). Expression of the Fanconi anemia gene FAC in human cell lines: lack of effect of oxygen tension. Blood Cells Mol Dis *21,* 182-91.

Johnstone, P., Reifsteck, C., Kohler, S., Worland, P., Olson, S. and Moses, R. E. (1997). Fanconi anemia group A and D cell lines respond normally to inhibitors of cell cycle regulation. Somat Cell Mol Genet *23,* 371-7.

Kaiser, T. N., Lojewski, A., Dougherty, C., Juergens, L., Sahar, E. and Latt, S. A. (1982). Flow cytometric characterization of the response of Fanconi's anemia cells to mitomycin C treatment. Cytometry *2,* 291-7.

Kim, U. J., Birren, B. W., Slepak, T., Mancino, V., Boysen, C., Kang, H. L., Simon, M. I. and Shizuya, H. (1996). Construction and characterization of a human bacterial artificial chromosome library. Genomics *34,* 213-8.

Kohli-Kumar, M., Morris, C., DeLaat, C., Sambrano, J., Masterson, M., Mueller, R., Shahidi, N. T., Yanik, G., Desantes, K., Friedman, D. J. and et, a. (1994). Bone marrow transplantation in Fanconi anemia using matched sibling donors. Blood *84,* 2050-4.

Korkina, L. G., Samochatova, E. V., Maschan, A. A., Suslova, T. B., Cheremisina, Z. P. and Afanas'ev, I. B. (1992). Release of active oxygen radicals by leukocytes of Fanconi anemia patients. J Leukoc Biol *52,* 357-62.

Kozel, P. J., Friedman, R. A., Erway, L. C., Yamoah, E. N., Liu, L. H., Riddle, T., Duffy, J. J., Doetschman, T., Miller, M. L., Cardell, E. L. and Shull, G. E. (1998). Balance and hearing deficits in mice with a null mutation in the gene encoding plasma membrane Ca2+-ATPase isoform 2. J Biol Chem *273,* 18693-6.

Krasnoshtein, F. and Buchwald, M. (1996). Developmental expression of the Fac gene correlates with congenital defects in Fanconi anemia patients. Hum Mol Genet *5,* 85-93.

Kruyt, F. A., Dijkmans, L. M., van-den-Berg, T. K. and Joenje, H. (1996). Fanconi anemia genes act to suppress a cross-linker-inducible p53-independent apoptosis pathway in lymphoblastoid cell lines. Blood *87,* 938-48.

Kruyt, F. A. and Youssoufian, H. (1998). The Fanconi anemia proteins FAA and FAC function in different cellular compartments to protect against cross-linking agent cytotoxicity. Blood *92,* 2229-36.

Kubbies, M., Schindler, D., Hoehn, H., Schinzel, A. and Rabinovitch, P. S. (1985). Endogenous blockage and delay of the chromosome cycle despite normal recruitment and growth phase explain poor proliferation and frequent edomitosis in Fanconi anemia cells. Am J Hum Genet *37,* 1022-30.

Kupfer, G. M. and D'Andrea, A. D. (1996). The effect of the Fanconi anemia polypeptide, FAC, upon p53 induction and G2 checkpoint regulation. Blood *88,* 1019-25.

Kupfer, G. M., Naf, D., Suliman, A., Pulsipher, M. and D'Andrea, A. D. (1997a). The Fanconi anaemia proteins, FAA and FAC, interact to form a nuclear complex. Nat Genet *17,* 487-90.

Kupfer, G. M., Yamashita, T., Naf, D., Suliman, A., Asano, S. and D'Andrea, A. D. (1997b). The Fanconi anemia polypeptide, FAC, binds to the cyclin-dependent kinase, cdc2. Blood *90,* 1047-54.

Kurtzberg, J. (1996). State of the art in umbilical cord transplantation. Oncology (Huntingt) *10,* 1086, 1091.

Kurtzberg, J., Laughlin, M., Graham, M. L., Smith, C., Olson, J. F., Halperin, E. C., Ciocci, G., Carrier, C., Stevens, C. E. and Rubinstein, P. (1996). Placental blood as a

source of hematopoietic stem cells for transplantation into unrelated recipients [see comments]. N Engl J Med *335,* 157-66.

Kuzmin, I., Stackhouse, T., Latif, F., Duh, F. M., Geil, L., Gnarra, J., Yao, M., Orcutt, M. L., Li, H., Tory, K. and et al. (1994). One-megabase yeast artificial chromosome and 400-kilobase cosmid-phage contigs containing the von Hippel-Lindau tumor suppressor and Ca(2+)- transporting adenosine triphosphatase isoform 2 genes. Cancer Res *54,* 2486-91.

Lambert, M. W., Tsongalis, G. J., Lambert, W. C., Hang, B. and Parrish, D. D. (1992). Defective DNA endonuclease activities in Fanconi's anemia cells, complementation groups A and B. Mutat Res *273,* 57-71.

Lambert, M. W., Tsongalis, G. J., Lambert, W. C. and Parrish, D. D. (1997). Correction of the DNA repair defect in Fanconi anemia complementation groups A and D cells. Biochem Biophys Res Commun *230,* 587-91.

Lennon, G., Auffray, C., Polymeropoulos, M. and Soares, M. B. (1996). The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. Genomics *33,* 151-2.

Liu, J. M. (1998). Gene transfer for the eventual treatment of Fanconi's anemia. Semin Hematol *35,* 168-79.

Liu, J. M., Auerbach, A. D. and Young, N. S. (1991). Fanconi anemia presenting unexpectedly in an adult kindred with no dysmorphic features. Am J Med *91,* 555-7.

Liu, J. M., Buchwald, M., Walsh, C. E. and Young, N. S. (1994). Fanconi anemia and novel strategies for therapy. Blood *84,* 3995-4007.

Liu, N., Lamerdin, J. E., Tucker, J. D., Zhou, Z. Q., Walter, C. A., Albala, J. S., Busch, D. B. and Thompson, L. H. (1997a). The human XRCC9 gene corrects chromosomal instability and mutagen sensitivities in CHO UV40 cells. Proc Natl Acad Sci U S A *94,* 9232-7.

Liu, N., Lamerdin, J. E., Tucker, J. D., Zhou, Z. Q., Walter, C. A., Albala, J. S., Busch, D. B. and Thompson, L. H. (1997b). The human XRCC9 gene corrects chromosomal instability and mutagen sensitivities in CHO UV40 cells. Proc Natl Acad Sci U S A *94,* 9232-7.

Lo Ten Foe, J. R., Kwee, M. L., Rooimans, M. A., Oostra, A. B., Veerman, A. J., van Weel, M., Pauli, R. M., Shahidi, N. T., Dokal, I., Roberts, I., Altay, C., Gluckman, E., Gibson, R. A., Mathew, C. G., Arwert, F. and Joenje, H. (1997). Somatic mosaicism in Fanconi anemia: molecular basis and clinical significance. Eur J Hum Genet *5,* 137-48.

Lo Ten Foe, J. R., Rooimans, M. A., Bosnoyan-Collins, L., Alon, N., Wijker, M., Parker, L., Lightfoot, J., Carreau, M., Callen, D. F., Savoia, A., Cheng, N. C., van Berkel, C. G. M., Strunk, M. H. P., Gille, J. J. P., Pals, G., Kruyt, F. A. E., Pronk, J. C., Arwert, F., Buchwald, M. and Joenje, H. (1996a). Expression cloning of a cDNA for the major Fanconi anaemia gene, FAA. Nat. Genet. *14,* 320-23.

Lo Ten Foe, J. R., Rooimans, M. A., Joenje, H. and Arwert, F. (1996b). Novel frameshift mutation (1806insA) in exon 14 of the Fanconi anemia C gene, FAC. Hum Mutat *7,* 264-5.

Martin, N. J. and Steinberg, B. G. (1983). The dup(3)(p25->pter) syndrome: a case with holoprosencephaly. Am J Med Genet *14,* 767-72.

Mavelli, I., Ciriolo, M. R., Rotilio, G., De Sole, P., Castorino, M. and Stabile, A. (1982). Superoxide dismutase, glutathione peroxidase and catalase in oxidative hemolysis. A study of Fanconi's anemia erythrocytes. Biochem Biophys Res Commun *106,* 286-90.

Mercer, W. E. and Schlegel, R. A. (1979). Phytohemagglutinin enhancement of cell fusion reduces polyethylene glycol cytotoxicity. Exp Cell Res *120,* 417-21.

Meyn, M. S. (1995). Ataxia-telangiectasia and cellular responses to DNA damage. Cancer Res *55,* 5991-6001.

Miller, S. A., Dykes, D. D. and Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res *16,* 1215.

Mohsen, A. W., Anderson, B. D., Volchenboum, S. L., Battaile, K. P., Tiffany, K., Roberts, D., Kim, J. J. and Vockley, J. (1998). Characterization of molecular defects in isovaleryl-CoA dehydrogenase in patients with isovaleric acidemia. Biochemistry *37,* 10325-35.

Moustacchi, E., Papadopoulo, D., Diatloff-Zito, C. and Buchwald, M. (1987). Two complementation groups of Fanconi's anemia differ in their phenotypic response to a DNA-crosslinking treatment. Hum Genet *75,* 45-47.

Naf, D., Kupfer, G. M., Suliman, A., Lambert, K. and D'Andrea, A. D. (1998). Functional activity of the fanconi anemia protein FAA requires FAC binding and nuclear localization. Mol Cell Biol *18,* 5952-60.

Nara N., M. T., et al (1980). Two siblings with Fanconi's Anemia developing squamous cell carcinomas. Rinsho Ketsueki *21,* 1944.

Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J. J. and de Jong, P. J. (1998). An improved approach for construction of bacterial artificial chromosome libraries [In Process Citation]. Genomics *52,* 1-8.

Papadopoulo, D., Guillouf, C., Mohrenweiser, H. and Moustacchi, E. (1990a). Hypomutability in Fanconi anemia cells is associated with increased deletion frequency at the HPRT locus. Proc Natl Acad Sci U S A *87,* 8383-7.

Papadopoulo, D., Laquerbe, A., Guillouf, C. and Moustacchi, E. (1993). Molecular spectrum of mutations induced at the HPRT locus by a cross-linking agent in human cell lines with different repair capacities. Mutat Res *294,* 167-77.

Papadopoulo, D., Porfirio, B. and Moustacchi, E. (1990b). Mutagenic response of Fanconi's anemia cells from a defined complementation group after treatment with photoactivated bifunctional psoralens. Cancer Res *50,* 3289-94.

Parker, L., dos Santos, C. and Buchwald, M. (1998). The delta327 mutation in the Fanconi anemia group C gene generates a novel transcript lacking the first two coding exons [In Process Citation]. Hum Mutat *Suppl,* S275-7.

Phipps, M. E., Latif, F., Prowse, A., Payne, S. J., Dietz-Band, J., Leversha, M., Affara, N. A., Moore, A. T., Tolmie, J., Schinzel, A. and et al. (1994). Molecular genetic analysis of the 3p- syndrome. Hum Mol Genet *3,* 903-8.

Pritsos, C. A. and Sartorelli, A. C. (1986). Generation of reactive oxygen radicals through bioactivation of mitomycin antibiotics. Cancer Res *46,* 3528-32.

Pronk, J. C., Gibson, R. A., Savoia, A., Wijker, M., Morgan, N. V., Melchionda, S., Ford, D., Temtamy, S., Ortega, J. J., Jansen, S. and et, a. (1995). Localisation of the Fanconi anaemia complementation group A gene to chromosome 16q24.3. Nat Genet *11,* 338-40.

Rackoff, W. R., Orazi, A., Robinson, C. A., Cooper, R. J., Alter, B. P., Freedman, M. H., Harris, R. E. and Williams, D. A. (1996). Prolonged administration of granulocyte colony-stimulating factor (filgrastim) to patients with Fanconi anemia: a pilot study. Blood *88,* 1588-93.

Rathbun, R. K., Faulkner, G. R., Ostroski, M. H., Christianson, T. A., Hughes, G., Jones, G., Cahn, R., Maziarz, R., Royle, G., Keeble, W., Heinrich, M. C., Grompe, M., Tower, P. A. and Bagby, G. C. (1997). Inactivation of the Fanconi anemia group C gene augments interferon-gamma-induced apoptotic responses in hematopoietic cells. Blood *90,* 974-85.

Rey, J. P., Scott, R. and Muller, H. (1994). Apoptosis is not involved in the hypersensitivity of Fanconi anemia cells to mitomycin C. Cancer Genet Cytogenet *75,* 67-71.

Ridet, A., Guillouf, C., Duchaud, E., Cundari, E., Fiore, M., Moustacchi, E. and Rosselli, F. (1997). Deregulated apoptosis is a hallmark of the Fanconi anemia syndrome. Cancer Res *57,* 1722-30.

Rosendorff, J., Bernstein, R., Macdougall, L. and Jenkins, T. (1987). Fanconi anemia: another disease of unusually high prevalence in the Afrikaans population of South Africa. Am J Med Genet *27,* 793-7.

Rosselli, F. and Moustacchi, E. (1990). Cocultivation of Fanconi anemia cells and of mouse lymphoma mutants leads to interspecies complementation of chromosomal hypersensitivity to DNA cross-linking agents. Hum Genet *84,* 517-21.

Rosselli, F., Ridet, A., Soussi, T., Duchaud, E., Alapetite, C. and Moustacchi, E. (1995). p53-dependent pathway of radio-induced apoptosis is altered in Fanconi anemia. Oncogene *10,* 9-17.

Rosselli, F., Sanceau, J., Gluckman, E., Wietzerbin, J. and Moustacchi, E. (1994). Abnormal lymphokine production: a novel feature of the genetic disease Fanconi anemia. II. In vitro and in vivo spontaneous overproduction of tumor necrosis factor alpha. Blood *83,* 1216-25.

Rosselli, F., Sanceau, J., Wietzerbin, J. and Moustacchi, E. (1992). Abnormal lymphokine production: a novel feature of the genetic disease Fanconi anemia. I. Involvement of interleukin-6. Hum Genet *89,* 42-8.

Ruppitsch, W., Meisslitzer, C., Weirich-Schwaiger, H., Klocker, H., Scheidereit, C., Schweiger, M. and Hirsch-Kauffmann, M. (1997). The role of oxygen metabolism for the pathological phenotype of Fanconi anemia. Hum Genet *99,* 710-9.

Saito, H., Hammond, A. T. and Moses, R. E. (1993). Hypersensitivity to oxygen is a uniform and secondary defect in Fanconi anemia cells. Mutat Res *294,* 255-62.

Saito, H. and Moses, R. E. (1992). Cytotoxicity of DNA crosslinking damage and oxygen tension in Fanconi anemia fibroblasts. Am J Hum Genet (Suppl.) *51,* 890.

Sala-Trepat, M., Boyse, J., Richard, P., Papadopoulo, D. and Moustacchi, E. (1993). Frequencies of HPRT- lymphocytes and glycophorin A variants erythrocytes in Fanconi anemia patients, their parents and control donors. Mutat Res *289,* 115-26.

Sasaki, M. S. and Tonomura, A. (1973). A high susceptibility of Fanconi's anemia to chromosome breakage by DNA cross-linking agents. Cancer Res *33,* 1829-36.

Savitsky, K., Bar-Shira, A., Gilad, S., Rotman, G., Ziv, Y., Vanagaite, L., Tagle, D. A., Smith, S., Uziel, T., Sfez, S. and et, a. (1995). A single ataxia telangiectasia gene with a product similar to PI-3 kinase. Science *268,* 1749-53.

Savoia, A., Centra, M., Ianzano, L., de-Cillis, G. P., Zelante, L. and Buchwald, M. (1995). Characterization of the 5' region of the Fanconi anaemia group C (FACC) gene. Hum Mol Genet *4,* 1321-6.

Schindler, D. and Hoehn, H. (1988). Fanconi anemia mutation causes cellular susceptibility to ambient oxygen. Am J Hum Genet *43,* 429-35.

Schmidt, H. H., Pirc-Danoewinata, H., Panzer-Grumayer, E. R., Sill, H., Sedlmayr, P., Neumeister, P., Linkesch, W. and Haas, O. A. (1998). Translocation (3;5)(p26;q13) in a patient with chronic T-cell lymphoproliferative disorder. Cancer Genet Cytogenet *104,* 82-5.

Schroeder, T. M., Anschutz, F. and Knopp, A. (1964). Spontaneous chromosome aberrations in familial panmyelopathy. Humangenetik *1,* 194-6.

Seyschab, H., Friedl, R., Sun, Y., Schindler, D., Hoehn-H, H. S. and Schroeder-Kurth, T. (1995). Comparative evaluation of diepoxybutane sensitivity and cellcycle blockage in the diagnosis of Fanconi anemia. Blood *85,* 2233-7.

Seyschab, H., Sun, Y., Friedl, R., Schindler, D. and Hoehn, H. (1993). G2 phase cell cycle disturbance as a manifestation of genetic cell damage. Hum Genet *92,* 61-8.

Shapiro, M. B. and Senapathy, P. (1987). RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res *15,* 7155-74.

Shizuya, H., Birren, B., Kim, U. J., Mancino, V., Slepak, T., Tachiiri, Y. and Simon, M. (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. Proc Natl Acad Sci U S A *89,* 8794-7.

Soci'e, G., Gluckman, E., Raynal, B., Petit, T., Landman-J, D. A. and Brison, O. (1993). Bone marrow transplantation for Fanconi anemia using low-dose cyclophosphamide/thoracoabdominal irradiation as conditioning regimen: chimerism study by the polymerase chain reaction. Blood *82,* 2249-56.

Somers, G. R., Tabrizi, S. N., Tiedemann, K., Chow, C. W., Garland, S. M. and Venter, D. J. (1995). Squamous cell carcinoma of the tongue in a child with Fanconi anemia: a case report and review of the literature. Pediatr Pathol Lab Med *15,* 597-607.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol *98,* 503-517.

Strathdee, C. A., Duncan, A. M. and Buchwald, M. (1992a). Evidence for at least four Fanconi anaemia genes including FACC on chromosome 9. Nat Genet *1,* 196-8.

Strathdee, C. A., Gavish, H., Shannon, W. R. and Buchwald, M. (1992b). Cloning of cDNAs for Fanconi's anemia by functional complementation. Nature *356,* 763-767.

Street, V. A., McKee-Johnson, J. W., Fonseca, R. C., Tempel, B. L. and Noben-Trauth, K. (1998). Mutations in a plasma membrane Ca2+-ATPase gene cause deafness in deafwaddler mice. Nat Genet *19,* 390-4.

Swift, M., Zimmerman, D. and McDonough, E. R. (1971). Squamous cell carcinomas in Fanconi's anemia. Jama *216,* 325-6.

Takeuchi, T. and Morimoto, K. (1993). Increased formation of 8-hydroxydeoxyguanosine, an oxidative DNA damage, in lymphoblasts from Fanconi's anemia patients due to possible catalase deficiency. Carcinogenesis *14,* 1115-20.

Thyagarajan, B. and Campbell, C. (1997). Elevated homologous recombination activity in fanconi anemia fibroblasts. J Biol Chem *272,* 23328-33.

Trask, B. J., Massa, H., Kenwrick, S. and Gitschier, J. (1991). Mapping of human chromosome Xq28 by two-color fluorescence in situ hybridization of DNA sequences to interphase cell nuclei. Am J Hum Genet *48,* 1-15.

van den Hoff, M. J. B., Moorman, A. F. M. and Lamers, W. H. (1992). Electroporation in 'intracellular' buffer increases cell survival. Nucl Acids Res *20,* 2902.

Verlander, P. C., Kaporis, A., Liu, Q., Zhang, Q., Seligsohn, U. and Auerbach, A. D. (1995). Carrier frequency of the IVS4 + 4 A-->T mutation of the Fanconi anemia gene FAC in the Ashkenazi Jewish population. Blood *86,* 4034-8.

Verlander, P. C., Lin, J. D., Udono, M. U., Zhang, Q., Gibson, R. A., Mathew, C. G. and Auerbach, A. D. (1994). Mutation analysis of the Fanconi anemia gene FACC. Am J Hum Genet *54,* 595-601.

Wagner, J. E. (1993). Umbilical cord blood stem cell transplantation. Am J Pediatr Hematol Oncol *15,* 169-74.

Walsh, C. E., Grompe, M., Vanin, E., Buchwald, M., Young, N. S., Nienhuis, A. W. and Liu, J. M. (1994a). A functionally active retrovirus vector for gene therapy in Fanconi anemia group C. Blood *84,* 453-9.

Walsh, C. E., Nienhuis, A. W., Samulski, R. J., Brown, M. G., Miller, J. L., Young, N. S. and Liu, J. M. (1994b). Phenotypic correction of Fanconi anemia in human hematopoietic cells with a recombinant adeno-associated virus vector. J Clin Invest *94,* 1440-8.

Wang, J., Otsuki, T., Youssoufian, H., Foe, J. L., Kim, S., Devetten, M., Yu, J., Li, Y., Dunn, D. and Liu, J. M. (1998). Overexpression of the fanconi anemia group C gene (FAC) protects hematopoietic progenitors from death induced by Fas-mediated apoptosis. Cancer Res *58,* 3538-41.

Weinert, T. A., Kiser, G. L. and Hartwell, L. H. (1994). Mitotic checkpoint genes in budding yeast and the dependence of mitosis on DNA replication and repair. Genes Dev *8,* 652-65.

Weksberg, R., Buchwald, M., Sargent, P. and Thompson-M-W, S. L. (1979). Specific cellular defects in patients with Fanconi anemia. J Cell Physiol *101,* 311-23.

Wevrick, R., Clarke, C. A. and Buchwald, M. (1993). Cloning and analysis of the murine Fanconi anemia group C cDNA. Human Molecular Genetics *2,* 655-662.

White, K., Grether, M. E., Abrams, J. M., Young, L., Farrell, K. and Steller, H. (1994). Genetic control of programmed cell death in Drosophila [see comments]. Science *264,* 677-83.

Whitney, M., Thayer, M., Reifsteck, C., Olson, S., Smith, L., Jakobs, P. M., Leach, R., Naylor, S., Joenje, H. and Grompe, M. (1995). Microcell mediated chromosome transfer maps the Fanconi anaemia group D gene to chromosome 3p. Nat Genet *11,* 341-3.

Whitney, M. A., Jakobs, P., Kaback, M., Moses, R. E. and Grompe, M. (1994). The Ashkenazi Jewish Fanconi anemia mutation: incidence among patients and carrier frequency in the at-risk population. Hum Mutat *3,* 339-41.

Whitney, M. A., Royle, G., Low, M. J., Kelly, M. A., Axthelm, M. K., Reifsteck, C., Olson, S., Braun, R. E., Heinrich, M. C., Rathbun, R. K., Bagby, G. C. and Grompe, M. (1996). Germ cell defects and hematopoietic hypersensitivity to gamma-interferon in mice with a targeted disruption of the Fanconi anemia C gene. Blood *88,* 49-58.

Whitney, M. A., Saito, H., Jakobs, P. M., Gibson, R. A., Moses, R. E. and Grompe, M. (1993). A common mutation in the FACC gene causes Fanconi anaemia in Ashkenazi Jews. Nature Genetics *4,* 202-205.

Willingale-Theune, J., Schweiger, M., Hirsch-Kauffmann, M., Meek, A. E., Paulin-Levasseur, M. and Traub, P. (1989). Ultrastructure of Fanconi anemia fibroblasts. J Cell Sci *93,* 651-65.

Wu, D. Y., Nozari, G., Schold, M., Conner, B. J. and Wallace, R. B. (1989). Direct analysis of single nucleotide variation in human DNA and RNA using in situ dot hybridization. Dna *8,* 135-42.

Yamashita, T., Barber, D. L., Zhu, Y., Wu, N. and D'Andrea, A. D. (1994). The Fanconi anemia polypeptide FACC is localized to the cytoplasm. Proc Natl Acad Sci U S A *91,* 6712-6.

Yamashita, T., Kupfer, G. M., Naf, D., Suliman, A., Joenje, H., Asano, S. and D'Andrea, A. D. (1998). The fanconi anemia pathway requires FAA phosphorylation and FAA/FAC nuclear accumulation. Proc Natl Acad Sci U S A *95,* 13085-90.

Young, N. and Alter, B. (1994). Aplastic Anemia, Aquired and Inherited, (Philadelphia: Saunders Company)

Youssoufian, H. (1994). Localization of Fanconi anemia C protein to the cytoplasm of mammalian cells. Proc Natl Acad Sci U S A *91,* 7975-9.

Youssoufian, H. (1995). Identification of cytosolic proteins that bind to the Fanconi anemia complementation group C polypeptide in vitro. Evidence for a multimeric complex
Localization of Fanconi anemia C protein to the cytoplasm of mammalian cells. J Biol Chem *270,* 9876-82.

Youssoufian, H. (1996). Cytoplasmic localization of FAC is essential for the correction of a prerepair defect in Fanconi anemia group C cells. J Clin Invest *97,* 2003-10.

Appendix A

Primers and Conditions for BAC End-sequences

BAC	Primers	Primer Sequence (5'-3')	PCR Conditions	Size
300K6	MG316 MG317	GACATTCTTCCAACCACTTTCC GGTGGATCAGGCTAGAAAAGG	46°C, 25s 72°C, 40s 94°C, 25s 37 cycles	303 bp
234M14	234F-1 234F-2	CAGACATCACCCCATCACTG GAGCCTTCATCTTTGGGTAGG	47°C, 25s 72°C, 30s 94°C, 25s 37 cycles	165 bp
187A24	187R-1 187R-2	CAAGTGAATCTTGAGCAACTGC TCTGTTCACTGACCACAGCTG	47°C, 25s 72°C, 25s 94°C, 25s 37 cycles	124 bp
187A24	187F-1 187F-2	TCCAGTGAACAAACATAGCACC TTTTAGCACTGGTGGGCAG	52°C, 25s 72°C, 25s 94°C, 25s 37 cycles	114 bp
183L15	183R-1 183R-2	AACCTTTATGGGAACCAATGC TTACAAAAGTGTGGGAGCCC	50°C, 25s 72°C, 30s 94°C, 25s 37 cycles	139 bp
293C22	293C22F-1 293C22F-2	TTACCTTCTGCCAAAATGGC TGCTCCTTGACTTATGATGGG	50°C, 25s 72°C, 30s 94°C, 25s 37 cycles	153 bp
20A10	20A10R-1 20A10R-2	ACTTAGGCAGTGTTTCCCTCC CCAAAAGTGTGAGGGAGGAC	44°C, 25s 72°C, 30s 94°C, 25s 37 cycles	218 bp
234M14	234R-3 234R-4	GAGGAAATGCATCCTTCTGC GCTGAAGGGGTTAGTTCGC	48°C, 25s 72°C, 25s 94°C, 25s 37 cycles	101 bp

123H15	123F-1 123F-2	TGCAAAATGAGGACAACAAG CTCAAGGCTCCGTCAAACTC	50°C, 25s 72°C, 25s 94°C, 25s 37 cycles	177 bp
343N13	343R-1 343R-2	CCTACTCCCAACCCTGGAG TTGAGTCCCATCCTCCAAAC	52°C, 25s 72°C, 25s 94°C, 25s 37 cycles	153 bp
102F6	102R-1 102R-2	GCACATTAAAACTCACTCCTGC GAGATGGGGCACCAGAAAC	50°C, 25s 72°C, 25s 94°C, 25s 37 cycles	183 bp
168J6	168JF-1 168JF-2	ACTGTGCTCTCCCTCCCC GAAGAATTTGTTAGTGGCCTGG	54°C, 25s 72°C, 25s 94°C, 25s 37 cycles	178 bp
334F6	334F-1 334F-2	CCTCTGGTAGTCTGCTTGGC AGATGTGGTCAGGCTCGG	52°C, 25s 72°C, 45s 94°C, 25s 37 cycles	147 bp
334F6	334R-1 334R-2	TTTGTGCATTTGAGTATTTGGC AAGTGGGATGCAGACTGACC	52°C, 25s 72°C, 45s 94°C, 25s 37 cycles	392 bp
269B10	269R-1 269R-2	GTCACTTTGCATTCCCTACTCC GGTAAGGCAGAAAGGAGCG	52°C, 25s 72°C, 45s 94°C, 25s 37 cycles	174 bp
269B10	269F-1 269F-2	CCATATTTTTTGTGTTCACAAGG AGAACCCCTCTTACAAGTTCCC	43°C, 25s 72°C, 45s 94°C, 25s 37 cycles	165 bp
177N7	177F-1 177F-2	CAGGGTCATGATATTCAACTCC CAAGACTTTTCCAGGCAGTG	53°C, 25s 72°C, 25s 94°C, 25s 37 cycles	104 bp
177N7	177R-1 177R-2	TCATTTCAGACAAATCCATTGC GCTGGGATTATAGGCGTGAC	52°C, 25s 72°C, 25s 94°C, 25s 37 cycles	196 bp

159N23	159F-1 159F-2	GAGCCGAAAGTAAGGGCTG TTGAGGAAGGAGAGCTAAGTGG	52°C, 25s 72°C, 25s 94°C, 25s 37 cycles	261 bp
93A5	93-R1 93-R2	ACCATTGGATTGCACCTACCAC TGAGATGGATTTCCCAGCAGTC	50°C, 25s 72°C, 25s 94°C, 25s 37 cycles	220 bp
P1 39F12	P39T-1 P39T-2	TTCATATTCTTCCCCCTCCC CTAACCACAAAAACCTTCCAGG	52°C, 25s 72°C, 25s 94°C, 25s 37 cycles	136 bp
P1 91E6	P91S-1 P91S-2	GCATATGCTTCAGAATCCCTG AAAGTCCCATGGCTGTGAAG	54°C, 25s 72°C, 25s 94°C, 25s 37 cycles	127 bp
P1 56F11	P56T-1 P56T-2	TGCACAGGAATGCTTAGCAC GTTGAATAGTCTTCCTCCCACG	52°C, 25s 72°C, 25s 94°C, 25s 37 cycles	259 bp
P1 56F11	P56S-1 P56S-2	CTGAAGAAGATTGAAACCCACC CATGAGCCGTGATTGAACC	52°C, 25s 72°C, 25s 94°C, 25s 37 cycles	150 bp
61D11 61D11	61D11R-1 61D11R-2	GGTCACAGAAAGTTCTAAGTGG GCCCTACGTCCTTTATGG	48°C, 25s 72°C, 30s 94°C, 25s 37 cycles	95 bp
61D11 61D11	61D11T-1 61D11T-2	TGTTCCAGATGAATCCAAAGCC GGTAGCAGGAGGTTAAATGCACC	54°C, 25s 72°C, 30s 94°C, 25s 37 cycles	130 bp
183L15	183F-3 183F-4	CACTTTACACTTTCCTGGACTCCAC TTGGGTCTGCCACTTAAAAGAGAG	54°C, 25s 72°C, 30s 94°C, 25s 37 cycles	130 bp
183L15	183NS3 183NS4	CGGGTGTTCAAAAAATACCAGG ACAGGTGAGAAGTTTCTCCCCC	56°C, 25s 72°C, 30s 94°C, 25s 37 cycles	95 bp

P1 39F12	39T-3 39T-4	TGAAACAGAGCAATGCCTGTCC TGCAAGGCACTGAGGAGATGAG	56°C, 25s 72°C, 30s 94°C, 25s 37 cycles	300 bp
393J18	393F-1 393F-2	ACGGGCAACAGATAGAAATTC TGACACATAGATAAAAGTGCTTGAG	52°C, 25s 72°C, 25s 94°C, 25s 37 cycles	118 bp
393J18	393R-1 393R-2	GCCTCTAAGTGTTCAAGTGAAAGG GTGCTGGGATAACAAGCGTG	54°C, 25s 72°C, 25s 94°C, 25s 37 cycles	105 bp
558E6	558F-1 558F-2	TTATCACCCCACTTTATAGAGCAG TTTCACCAGGCGGATAGTTC	54°C, 25s 72°C, 25s 94°C, 25s 37 cycles	142 bp
187P1	187PF-1 187PF-2	CCCAGGTTATGGAACTCAGC TATGTGCTGCTCCATCCTTG	52°C, 25s 72°C, 25s 94°C, 25s 37 cycles	92 bp
187P1	187PR-1 187PR-2	CAACTTGTGTCTCAGTGCAGTTC GATTTCCAATACAGGCAGTGG	52°C, 25s 72°C, 25s 94°C, 25s 37 cycles	112 bp
109F16	109F-1 109F-2	CATGGCAGTCATTACTCTGTTTC ATCTGAAGGGGATGTTGAGG	56°C, 25s 72°C, 25s 94°C, 25 37 cycles	108 bp
109F16	109R-1 109R-2	CCATATTGGGCACTGACTTATC GACTAAAGCAGTTTTCAGGGC	56°C, 25s 72°C, 25s 94°C, 25s 37 cycles	146 bp
168D4	168DF-1 168DF-2	CTCCACTGGATGTAACAATGTTC CCTTCCCATCACTTTGGTGG	50°C, 25s 72°C, 30s 94°C, 25s 37 cycles	113 bp
168D4	168DR-1 168DR-2	CCAACCCACATCTCACAGGAG CCCTTATCCTGCCCTCAATG	50°C, 25s 72°C, 30s 94°C, 25s 37 cycles	151 bp

100I3	100F-1 100F-2	TTCTGGAAGTCAGGCTGGC TCATTTATTTCCAGCCCCC	52°C, 25s 72°C, 25s 94°C, 25s 37 cycles	99 bp
100I3	100R-1 100R-2	CCAAGGCTATCCAGTGAACAAAC AAAGAATCGGGGAAGAAGGAAG	50°C, 25s 72°C, 25s 94°C, 25s 37 cycles	181 bp
140G19	140F-1 140F-2	TCAGTGATCTCTGTCAGGGAGG CGAGGCTGAGGAATTGTTTTAC	52°C, 25s 72°C, 25s 94°C, 25s 37 cycles	140 bp
172I17 172I17	172F-1 172F-2	CCACAGCTTCATCAAGGGTG TGCAGTTATGGATGGAGGATG	52°C, 25s 72°C, 25s 94°C, 25s 37 cycles	110 bp
448N22	448F-1 448F-2	ACAGGCTAGATTTGCATGTTTC CGGTATAAAGGGAGAATGGC	52°C, 25s 72°C, 30s 94°C, 25s 37 cycles	100 bp
448N22	448R-1 448R-2	GACAACACTGAAAAGGGAACG CCACCATTTTGCATACGATG	50°C, 25s 72°C, 30s 94°C, 25s 37 cycles	120 bp
134A23	134F-1 134F-2	AGACCTCAGGACATCAGAGGC GCATGAGCAGATGGAGTATGG	52°C, 25s 72°C, 25s 94°C, 25s 37 cycles	105 bp
134A23	134R-1 134R-2	GTTGACCCCAAGGTGAAATG GACAATGCTAACGCCTTTGC	50°C, 25s 72°C, 25s 94°C, 25s 37 cycles	104 bp
320K16	320F-1 320F-2	GGTCTTTTTCAACTCTGTGGC AGCAATTTCAGCAGGTCAGC	50°C, 25s 72°C, 25s 94°C, 25s 37 cycles	156 bp
558E6	558R-1 558R-2	CCCAAGTTCCATGAAGTGTATG TGATGAGTACACAGGGTTCAGAG	52°C, 25s 72°C, 45s 94°C, 25s 37 cycles	441 bp

551O20	551R-1 551R-2	GCAAAGGAGTTTTCTGGGAGGC AAGTGTCAAGCAGGGCTTCTGGAC	50°C, 25s 72°C, 25s 94°C, 25s 37 cycles	118 bp
551O20	551F-1 551F-2	GTATGTCCATGATGAAAGCTGCTG AAATATGCCCCTGAGGTTTTCC	50°C, 25s 72°C, 25s 94°C, 25s 37 cycles	116 bp
152B1	152R-1 152R-2	GTTGCATACAGCAAGAGGATAGAGC CAACAAAGGACTGGTCAGGAAGC	50°C, 25s 72°C, 25s 94°C, 25s 37 cycles	246 bp
152B1 152B1	152F-1 152F-2	CCTTTCTCCTCTCCACCCTTAAAC CTTTTTCCAGCCTCCTGAAGAAC	50°C, 25s 72°C, 25s 94°C, 25s 37 cycles	187 bp

Appendix B

FANCD cDNA Sequence

```
-63                    TCGAAAACTACGGGCGGCGACGGCTTCTCGGAAGTAATTTAAGTGCACAAGACATTGGTCAAA -1

  1 ATG GTT TCC AAA AGA AGA CTG TCA AAA TCT GAG GAT AAA GAG AGC CTG ACA GAA GAT GCC  60
  1 M   V   S   K   R   R   L   S   K   S   E   D   K   E   S   L   T   E   D   A   20

 61 TCC AAA ACC AGG AAG CAA CCA CTT TCC AAA AAG ACA AAG AAA TCT CAT ATT GCT AAT GAA  120
 21 S   K   T   R   K   Q   P   L   S   K   K   T   K   K   S   H   I   A   N   E   40

121 GTT GAA GAA AAT GAC AGC ATC TTT GTA AAG CTT CTT AAG ATA TCA GGA ATT ATT CTT AAA  180
 41 V   E   E   N   D   S   I   F   V   K   L   L   K   I   S   G   I   I   L   K   60

181 ACG GGA GAG AGT CAG AAT CAA CTA GCT GTG GAT CAA ATA GCT TTC CAA AAG AAG CTC TTT  240
 61 T   G   E   S   Q   N   Q   L   A   V   D   Q   I   A   F   Q   K   K   L   F   80

241 CAG ACC CTG AGG AGA CAC CCT TCC TAT CCC AAA ATA ATA GAA GAA TTT GTT AGT GGC CTG  300
 81 Q   T   L   R   R   H   P   S   Y   P   K   I   I   E   E   F   V   S   G   L  100

301 GAG TCT TAC ATT GAG GAT GAA GAC AGT TTC AGG AAC TGC CTT TTG TCT TGT GAG CGT CTG  360
101 E   S   Y   I   E   D   E   D   S   F   R   N   C   L   L   S   C   E   R   L  120

361 CAG GAT GAG GAA GCC AGT ATG GGT GCA TCT TAT TCT AAG AGT CTC ATC AAA CTG CTT CTG  420
121 Q   D   E   E   A   S   M   G   A   S   Y   S   K   S   L   I   K   L   L   L   14

421 GGG ATT GAC ATA CTG CAG CCT GCC ATT ATC AAA ACC TTA TTT GAG AAG TTG CCA GAA TAT  480
141 G   I   D   I   L   Q   P   A   I   I   K   T   L   F   E   K   L   P   E   Y  160

481 TTT TTT GAA AAC AAG AAC AGT GAT GAA ATC AAC ATA CCT CGA CTC ATT GTC AGT CAA CTA  540
161 F   F   E   N   K   N   S   D   E   I   N   I   P   R   L   I   V   S   Q   L  180

541 AAA TGG CTT GAC AGA GTT GTG GAT GGC AAG GAC CTC ACC ACC AAG ATC ATG CAG CTG ATC  600
181 K   W   L   D   R   V   V   D   G   K   D   L   T   T   K   I   M   Q   L   I  200

601 AGT ATT GCT CCA GAG AAC CTG CAG CAT GAC ATC ATC ACC AGC CTA CCT GAG ATC CTA GGG  660
201 S   I   A   P   E   N   L   Q   H   D   I   I   T   S   L   P   E   I   L   G  220

661 GAT TCC CAG CAC GCT GAT GTG GGG AAA GAA CTC AGT GAC CTA CTG ATA GAG AAT ACT TCA  720
221 D   S   Q   H   A   D   V   G   K   E   L   S   D   L   L   I   E   N   T   S  240

721 CTC ACT GTC CCA ATC CTG GAT GTC CTT TCA AGC CTC CGA CTT GAC CCA AAC TTC CTA TTG  780
241 L   T   V   P   I   L   D   V   L   S   S   L   R   L   D   P   N   F   L   L  260

781 AAG GTT CGC CAG TTG GTG ATG GAT AAG TTG TCG TCT ATT AGA TTG GAG GAT TTA CCT GTG  840
261 K   V   R   Q   L   V   M   D   K   L   S   S   I   R   L   E   D   L   P   V  280

841 ATA ATA AAG TTC ATT CTT CAT TCC GTA ACA GCC ATG GAT ACA CTT GAG GTA ATT TCT GAG  900
281 I   I   K   F   I   L   H   S   V   T   A   M   D   T   L   E   V   I   S   E  300

901 CTT CGG GAG AAG TTG GAT CTG CAG CAT TGT GTT TTG CCA TCA CGG TTA CAG GCT TCC CAA  960
301 L   R   E   K   L   D   L   Q   H   C   V   L   P   S   R   L   Q   A   S   Q  320

961 GTA AAG TTG AAA AGT AAA GGA CGA GCA AGT TCC TCA GGA AAT CAA GAA AGC AGC GGT CAG 1020
321 V   K   L   K   S   K   G   R   A   S   S   S   G   N   Q   E   S   S   G   Q  340
```

```
1021 AGC TGT ATT ATT CTC CTC TTT GAT GTA ATA AAG TCA GCT ATT AGA TAT GAG AAA ACC ATT  1080
 341 S   C   I   I   L   L   F   D   V   I   K   S   A   I   R   Y   E   K   T   I    360

1081 TCA GAA GCC TGG ATT AAG GCA ATT GAA AAC ACT GCC TCA GTA TCT GAA CAC AAG GTG TTT  1140
 361 S   E   A   W   I   K   A   I   E   N   T   A   S   V   S   E   H   K   V   F    380

1141 GAC CTG GTG ATG CTT TTC ATC ATC TAT AGC ACC AAT ACT CAG ACA AAG AAG TAC ATT GAC  1200
 381 D   L   V   M   L   F   I   I   Y   S   T   N   T   Q   T   K   K   Y   I   D    400

1201 AGG GTG CTA AGA AAT AAG ATT CGA TCA GGC TGC ATT CAA GAA CAG CTG CTC CAG AGT ACA  1260
 401 R   V   L   R   N   K   I   R   S   G   C   I   Q   E   Q   L   L   Q   S   T    420

1261 TTC TCT GTT CAT TAC TTA GTT CTT AAG GAT ATG TGT TCA TCC ATT CTG TCG CTG GCT CAG  1320
 421 F   S   V   H   Y   L   V   L   K   D   M   C   S   S   I   L   S   L   A   Q    440

1321 AGT TTG CTT CAC TCT CTA GAC CAG AGT ATA ATT TCA TTT GGC AGT CTC CTA TAC AAA TAT  1380
 441 S   L   L   H   S   L   D   Q   S   I   I   S   F   G   S   L   L   Y   K   Y    460

1381 GCA TTT AAG TTT TTT GAC ACG TAC TGC CAG CAG GAA GTG GTT GGT GCC TTA GTG ACC CAT  1440
 461 A   F   K   F   F   D   T   Y   C   Q   Q   E   V   V   G   A   L   V   T   H    480

1441 ATC TGC AGT GGG AAT GAA GCT GAA GTT GAT ACT GCC TTA GAT GTC CTT CTA GAG TTG GTA  1500
 481 I   C   S   G   N   E   A   E   V   D   T   A   L   D   V   L   L   E   L   V    500

1501 GTG TTA AAC CCA TCT GCT ATG ATG ATG AAT GCT GTC TTT GTA AAG GGC ATT TTA GAT TAT  1560
 501 V   L   N   P   S   A   M   M   M   N   A   V   F   V   K   G   I   L   D   Y    520

1561 CTG GAT AAC ATA TCC CCT CAG CAA ATA CGA AAA CTC TTC TAT GTT CTC AGC ACA CTG GCA  1620
 521 L   D   N   I   S   P   Q   Q   I   R   K   L   F   Y   V   L   S   T   L   A    540

1621 TTT AGC AAA CAG AAT GAA GCC AGC AGC CAC ATC CAG GAT GAC ATG CAC TTG GTG ATA AGA  1680
 541 F   S   K   Q   N   E   A   S   S   H   I   Q   D   D   M   H   L   V   I   R    560

1681 AAG CAG CTC TCT AGC ACC GTA TTC AAG TAC AAG CTC ATT GGG ATT ATT GGT GCT GTG ACC  1740
 561 K   Q   L   S   S   T   V   F   K   Y   K   L   I   G   I   I   G   A   V   T    580

1741 ATG GCT GGC ATC ATG GCG GCA GAC AGA AGT GAA TCA CCT AGT TTG ACC CAA GAG AGA GCC  1800
 581 M   A   G   I   M   A   A   D   R   S   E   S   P   S   L   T   Q   E   R   A    600

1801 AAC CTG AGC GAT GAG CAG TGC ACA CAG GTG ACC TCC TTG TTG CAG TTG GTT CAT TCC TGC  1860
 601 N   L   S   D   E   Q   C   T   Q   V   T   S   L   L   Q   L   V   H   S   C    620

1861 AGT GAG CAG TCT CCT CAG GCC TCT GCA CTT TAC TAT GAT GAA TTT GCC AAC CTG ATC CAA  1920
 621 S   E   Q   S   P   Q   A   S   A   L   Y   Y   D   E   F   A   N   L   I   Q    640

1921 CAT GAA AAG CTG GAT CCA AAA GCC CTG GAA TGG GTT GGG CAT ACC ATC TGT AAT GAT TTC  1980
 641 H   E   K   L   D   P   K   A   L   E   W   V   G   H   T   I   C   N   D   F    660

1981 CAG GAT GCC TTC GTA GTG GAC TCC TGT GTT GTT CCG GAA GGT GAC TTT CCA TTT CCT GTG  2040
 661 Q   D   A   F   V   V   D   S   C   V   V   P   E   G   D   F   P   F   P   V    680

2041 AAA GCA CTG TAC GGA CTG GAA GAA TAC GAC ACT CAG GAT GGG ATT GCC ATA AAC CTC CTG  2100
 681 K   A   L   Y   G   L   E   E   Y   D   T   Q   D   G   I   A   I   N   L   L    700

2101 CCG CTG CTG TTT TCT CAG GAC TTT GCA AAA GAT GGG GGT CCG GTG ACC TCA CAG GAA TCA  2160
 701 P   L   L   F   S   Q   D   F   A   K   D   G   G   P   V   T   S   Q   E   S    720

2161 GGC CAA AAA TTG GTG TCT CCG CTG TGC CTG GCT CCG TAT TTC CGG TTA CTG AGA CTT TGT  2220
 721 G   Q   K   L   V   S   P   L   C   L   A   P   Y   F   R   L   L   R   L   C    740

2221 GTG GAG AGA CAG CAT AAC GGA AAC TTG GAG GAG ATT GAT GGT CTA CTA GAT TGT CCT ATA  2280
 741 V   E   R   Q   H   N   G   N   L   E   E   I   D   G   L   L   D   C   P   I    760
```

```
2281 TTC CTA ACT GAC CTG GAG CCT GGA GAG AAG TTG GAG TCC ATG TCT GCT AAA GAG CGT TCA  2340
 761 F   L   T   D   L   E   P   G   E   K   L   E   S   M   S   A   K   E   R   S    780

2341 TTC ATG TGT TCT CTC ATA TTT CTT ACT CTC AAC TGG TTC CGA GAG ATT GTA AAT GCC TTC  2400
 781 F   M   C   S   L   I   F   L   T   L   N   W   F   R   E   I   V   N   A   F    800

2401 TGC CAG GAA ACA TCA CCT GAG ATG AAG GGG AAG GTG CTC ACT CGG TTA AAG CAC ATT GTA  2460
 801 C   Q   E   T   S   P   E   M   K   G   K   V   L   T   R   L   K   H   I   V    820

2461 GAA TTG CAA ATA ATC CTG GAA AAG TAC TTG GCA GTC ACC CCA GAC TAT GTC CCT CCT CTT  2520
 821 E   L   Q   I   I   L   E   K   Y   L   A   V   T   P   D   Y   V   P   P   L    840

2521 GGA AAC TTT GAT GTG GAA ACT TTA GAT ATA ACA CCT CAT ACT GTT ACT GCT ATT TCA GCA  2580
 841 G   N   F   D   V   E   T   L   D   I   T   P   H   T   V   T   A   I   S   A    860

2581 AAA ATC AGA AAG AAA GGA AAA ATA GAA AGG AAA CAA AAA ACA GAT GGC AGC AAG ACA TCC  2640
 861 K   I   R   K   K   G   K   I   E   R   K   Q   K   T   D   G   S   K   T   S    880

2641 TCC TCT GAC ACA CTT TCA GAA GAG AAA AAT TCA GAA TGT GAC CCT ACG CCA TCT CAT AGA  2700
 881 S   S   D   T   L   S   E   E   K   N   S   E   C   D   P   T   P   S   H   R    900

2701 GGC CAG CTA AAC AAG GAG TTC ACA GGG AAG GAA GAA AAG ACA TCA TTG TTA CTA CAT AAT  2760
 901 G   Q   L   N   K   E   F   T   G   K   E   E   K   T   S   L   L   L   H   N    920

2761 TCC CAT GCT TTT TTC CGA GAG CTG GAC ATT GAG GTC TTC TCT ATT CTA CAT TGT GGA CTT  2820
 921 S   H   A   F   F   R   E   L   D   I   E   V   F   S   I   L   H   C   G   L    940

2821 GTG ACG AAG TTC ATC TTA GAT ACT GAA ATG CAC ACT GAA GCT ACA GAA GTT GTG CAA CTT  2880
 941 V   T   K   F   I   L   D   T   E   M   H   T   E   A   T   E   V   V   Q   L    960

2881 GGG CCC CCT GAG CTG CTT TTC TTG CTG GAA GAT CTC TCC CAG AAG CTG GAG AGT ATG CTG  2940
 961 G   P   P   E   L   L   F   L   L   E   D   L   S   Q   K   L   E   S   M   L    980

2941 ACA CCT CCT ATT GCC AGG AGA GTC CCC TTT CTC AAG AAC AAA GGA AGC CGG AAT ATT GGA  3000
 981 T   P   P   I   A   R   R   V   P   F   L   K   N   K   G   S   R   N   I   G   1000

3001 TTC TCA CAT CTC CAA CAG AGA TCT GCC CAA GAA ATT GTT CAT TGT GTT TTT CAA CTG CTG  3060
1001 F   S   H   L   Q   Q   R   S   A   Q   E   I   V   H   C   V   F   Q   L   L   1020

3061 ACC CCA ATG TGT AAC CAC CTG GAG AAC ATT CAC AAC TAT TTT CAG TGT TTA GCT GCT GAG  3120
1021 T   P   M   C   N   H   L   E   N   I   H   N   Y   F   Q   C   L   A   A   E   1040

3121 AAT CAC GGT GTA GTT GAT GGA CCA GGA GTG AAA GTT CAG GAG TAC CAC ATA ATG TCT TCC  3180
1041 N   H   G   V   V   D   G   P   G   V   K   V   Q   E   Y   H   I   M   S   S   1060

3181 TGC TAT CAG AGG CTG CTG CAG ATT TTT CAT GGG CTT TTT GCT TGG AGT GGA TTT TCT CAA  3240
1061 C   Y   Q   R   L   L   Q   I   F   H   G   L   F   A   W   S   G   F   S   Q   1080

3241 CCT GAA AAT CAG AAT TTA CTG TAT TCA GCC CTC CAT GTC CTT AGT AGC CGA CTG AAA CAG  3300
1081 P   E   N   Q   N   L   L   Y   S   A   L   H   V   L   S   S   R   L   K   Q   1100

3301 GGA GAA CAC AGC CAG CCT TTG GAG GAA CTA CTC AGC CAG AGC GTC CAT TAC TTG CAG AAT  3360
1101 G   E   H   S   Q   P   L   E   E   L   L   S   Q   S   V   H   Y   L   Q   N   1120

3361 TTC CAT CAA AGC ATT CCC AGT TTC CAG TGT GCT CTT TAT CTC ATC AGA CTT TTG ATG GTT  3420
1121 F   H   Q   S   I   P   S   F   Q   C   A   L   Y   L   I   R   L   L   M   V   1140

3421 ATT TTG GAG AAA TCA ACA GCT TCT GCT CAG AAC AAA GAA AAA ATT GCT TCC CTT GCC AGA  3480
1141 I   L   E   K   S   T   A   S   A   Q   N   K   E   K   I   A   S   L   A   R   1160

3481 CAA TTC CTC TGT CGG GTG TGG CCA AGT GGG GAT AAA GAG AAG AGC AAC ATC TCT AAT GAC  3540
1161 Q   F   L   C   R   V   W   P   S   G   D   K   E   K   S   N   I   S   N   D   1180
```

```
3541 CAG CTC CAT GCT CTG CTC TGT ATC TAC CTG GAG CAC ACA GAG AGC ATT CTG AAG GCC ATA  3600
1181 Q   L   H   A   L   L   C   I   Y   L   E   H   T   E   S   I   L   K   A   I    1200

3601 GAG GAG ATT GCT GGT GTT GGT GTC CCA GAA CTG ATC AAC TCT CCT AAA GAT GCA TCT TCC  3660
1201 E   E   I   A   G   V   G   V   P   E   L   I   N   S   P   K   D   A   S   S    1220

3661 TCC ACA TTC CCT ACA CTG ACC AGG CAT ACT TTT GTT GTT TTC TTC CGT GTG ATG ATG GCT  3720
1221 S   T   F   P   T   L   T   R   H   T   F   V   V   F   F   R   V   M   M   A    1240

3721 GAA CTA GAG AAG ACG GTG AAA AAA ATT GAG CCT GGC ACA GCA GCA GAC TCG CAG CAG ATT  3780
1241 E   L   E   K   T   V   K   K   I   E   P   G   T   A   A   D   S   Q   Q   I    1260

3781 CAT GAA GAG AAA CTC CTC TAC TGG AAC ATG GCT GTT CGA GAC TTC AGT ATC CTC ATC AAC  3840
1261 H   E   E   K   L   L   Y   W   N   M   A   V   R   D   F   S   I   L   I   N    1280

3841 TTG ATA AAG GTA TTT GAT AGT CAT CCT GTT CTG CAT GTA TGT TTG AAG TAT GGG CGT CTC  3900
1281 L   I   K   V   F   D   S   H   P   V   L   H   V   C   L   K   Y   G   R   L    1300

3901 TTT GTG GAA GCA TTT CTG AAG CAA TGT ATG CCG CTC CTA GAC TTC AGT TTT AGA AAA CAC  3960
1301 F   V   E   A   F   L   K   Q   C   M   P   L   L   D   F   S   F   R   K   H    1320

3961 CGG GAA GAT GTT CTG AGC TTA CTG GAA ACC TTC CAG TTG GAC ACA AGG CTG CTT CAT CAC  4020
1321 R   E   D   V   L   S   L   L   E   T   F   Q   L   D   T   R   L   L   H   H    1340

4021 CTG TGT GGG CAT TCC AAG ATT CAC CAG GAC ACG AGA CTC ACC CAA CAT GTG CCT CTG CTC  4080
1341 L   C   G   H   S   K   I   H   Q   D   T   R   L   T   Q   H   V   P   L   L    1360

4081 AAA AAG ACC CTG GAA CTT TTA GTT TGC AGA GTC AAA GCT ATG CTC ACT CTC AAC AAT TGT  4140
1361 K   K   T   L   E   L   L   V   C   R   V   K   A   M   L   T   L   N   N   C    1380

4141 AGA GAG GCT TTC TGG CTG GGC AAT CTA AAA AAC CGG GAC TTG CAG GGT GAA GAG ATT AAG  4200
1381 R   E   A   F   W   L   G   N   L   K   N   R   D   L   Q   G   E   E   I   K    1400

4201 TCC CAA AAT TCC CAG GAG AGC ACA GCA GAT GAG AGT GAG GAT GAC ATG TCA TCC CAG GCC  4260
1401 S   Q   N   S   Q   E   S   T   A   D   E   S   E   D   D   M   S   S   Q   A    1420

4261 TCC AAG AGC AAA GCC ACT GAG GTA TCT CTA CAA AAC CCA CCA GAG TCT GGC ACT GAT GGT  4320
1421 S   K   S   K   A   T   E   V   S   L   Q   N   P   P   E   S   G   T   D   G    1440

4321 TGC ATT TTG TTA ATT GTT CTA AGT TGG TGG AGC AGA ACT TTG CCT ACT TAT GTT TAT TGT  4380
1441 C   I   L   L   I   V   L   S   W   W   S   R   T   L   P   T   Y   V   Y   C    1460

4381 CAA ATG CTT CTA TGC CCA TTT CCA TTC CCT CCA TAA CAGCTTCTGTGCTTATATAATTTTTGGGACC  4447
1461 Q   M   L   L   C   P   F   P   F   P   P   *                                    1472

4448 CAGAAGAAACAACGACACAATCTTAGAATCACTCCTGAGTATCTCGAGTTGTGGCATTTGTTATAGAGTTGACAATTTT  4526

4527 CTGCATTATAGCCTCTCATTTTCCATGAATTCATATCTGAAACCATTTTAGAAGGGAGAAGTCATCGAAGTATTTTCTG  4605

4606 AGTGTTGAGAAGAATGAGTTAAACCATTTAAACACATTTGAAACATACAAAAATAGAAATGTGAAAGCATTTGGTGAAA  4684

4685 GCCAAAGCACAGAGTCAGAAGCTGCCACCTTAGAGAACTGAAATAAAAATAGAAGTTCTTACGCTTTTTTGTGGTACAG  4763

4764 ATGCTTTCGACAATTTAAAGAAAGCTAAATAAAAATGTAGACATGGCTGGCGCAGTGGCTCATGCTTGTAATCCTAGCA  4842

4843 CTTTTTGAGGCCAAGGTAGGAGGATTGCTTGAGTCCGGGAGCTCAAGGCAAAGCTGCACAACATAACAAGACCCTATCT  4921

4922 CCACAAAAAAAAATGAAAAATAAACCTGGGTGCGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCGATGTGGGC  5000

5001 AGATCACAAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATAGTGAAACCCCATCTCTACTGAAAATACAAAAATTAGC  5079

5080 TGGGTGTGGTGGCACGTGCCTGTTATCTCAGCTACTTGGGAAGCTGA                                  5127
```