AD_____

Award Number: DAMD17-00-1-0538

TITLE: DNA Repair and Checkpoint Genes as NF1 Modifiers

PRINCIPAL INVESTIGATOR: Andre Bernards, Ph.D.

CONTRACTING ORGANIZATION: The General Hospital Corporation
Boston, Massachusetts 02114-2554

REPORT DATE: November 2003

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	November 2003	Final (1 Oct 2000 - 1 Nov 2003)

4. TITLE AND SUBTITLE
DNA Repair and Checkpoint Genes as NF1 Modifiers

5. FUNDING NUMBERS
DAMD17-00-1-0538

6. AUTHOR(S)
Andre Bernards, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The General Hospital Corporation
Boston, Massachusetts 02114-2554

E-Mail: abernard@helix.mgh.harvard.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

This study is aimed to determine whether common protein altering SNP alleles of DNA repair or DNA damage-associated checkpoint genes are associated with higher or lower than average neurofibroma burden in NF1 patients. As part of this project, we identified 793 missense single nucleotide polymorphisms (SNPs) in 293 candidate modifier genes. We also generated three relational databases to manage SNP and genotype information. Beyond data mining and generating information handling tools, we recruited approximately 80 eligible patients using our originally planned recruitment strategy. Because recruitment fell short of the required 600 patients, during the final year of this grant, we enlisted six additional clinical collaborators. With recruitment continuing, we evaluated several high throughput genotyping methods. Single base extension fluorescence polarization genotyping was deemed too cumbersome, but using allele-specific PCR or restriction fragment length polymorphism genotyping, we determined ~20,000 individual genotypes for 37 SNPs in 26 genes. Three grant proposals have been submitted based on preliminary data obtained in this project, and NIH R01 and Army Investigator-Initiated Research Grants have recently been awarded.

14. SUBJECT TERMS
No Subject Terms Provided.

15. NUMBER OF PAGES
14

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

Introduction

Neurofibromatosis type 1 (NF1) is a common genetic disorder that affects 2 to 3 per 10,000 worldwide. Patients are at increased risk of developing a diverse set of symptoms, the most common of which include skin pigmentation defects, iris hamartomas, benign tumors associated with the peripheral nervous system, termed neurofibromas, and learning problems (Huson and Hughes, 1994). NF1 is paradigmatic for a disease with variable expressivity and genetic studies have implicated modifier genes as important determinants of symptomatic variability, and most notably of variability in neurofibroma burden (Easton et al., 1993; Szudek et al., 2000). This project aimed to create the informatics and genetic resources to identify modifiers of neurofibroma burden and to explore whether genes involved in maintaining genome stability play rate limiting roles in neurofibroma development. We focus on genes that modify neurofibroma development, because these benign tumors contribute significantly to the overall morbidity of NF1 and because their numerical variability is a cause for significant patient anxiety as well as a major problem for clinical trials.

Body

The Statement of Work listed as **Task 1** the creation of computerized patient and modifier gene databases. This task was accomplished as planned during the first month of funding, but we have continued to modify and expand the single nucleotide polymorphism (SNP) database far beyond what we had envisaged for the entire funding period. The password-protected patient database includes names, sex, dates of birth, information on neurofibroma numbers, information on potential confounders (number of pregnancies and information on whether disease is familial or sporadic), contact information, details about consent procedures, summaries of correspondence, codes used to identify samples in the laboratory, and other information if available. It currently contains information on 296 NF1 patients. Of these patients, 21 were seen at the MGH NF clinic by our collaborator Dr. Mia MacCollin. An additional 28 were brought to our attention by collaborating with Dr Andreas Kurtz, as suggested by the integration panel. The remaining patients contacted the Principal Investigator or the project associated Genetic Counselor after leaning about this study, mostly from notices posted by patient organizations. No new patient information has been entered into this database for the past several months, because new HIPAA requirements no

longer make it possible to collect patient-related information without first obtaining signed waivers. Thus, the actual number of patients who contacted us is about 5% higher than the number indicated above.

Our original proposal was to perform a case-control allele association study among initially 300 and eventually 600 eligible NF1 patients. Eligibility criteria were designed to select for patients that represent the top and bottom 20% of neurofibroma burden in various age groups (Table 1)

Age	Number of Neurofibromas
18-20 years old	fewer than 5 or more than 30
20-30 years old	fewer than 10 or more than 100
30-40 years old	fewer than 20 or more than 200
40-50 years old	more than 500
40 years or older	fewer than 50

Table 1: Study eligibility criteria. Neurofibromas are benign tumors that can be pinched off with the skin. Any tumor is counted, regardless of its size.

We proposed to genotype common protein altering SNP alleles of candidate modifier genes identified in a screen performed by collaborators at the MIT Center for Genome Research. In reality the MIT screen only included a small number of the potential modifier genes that were of interest to us. Thus, rather than limiting ourselves to just the few genes analyzed at MIT, we invested an estimated 500-750 hours in literature and online database surveys, and in other "data mining" efforts in order to identify candidate modifier alleles among a comprehensive set of genes implicated in maintaining genome stability. This far more ambitious approach was made possible by the identification of well over one million SNPs during the early phases of the human genome project. Mining of online SNP and literature databases during the first year of funding identified 325 protein altering SNPs in 185 potential neurofibroma modifiers. 57 if these missense SNPs (17.5%) had a minor allele frequency >4%. Continued data mining has presently identified 793 nonsynonymous alleles of 293 potential modifier genes, 155 (19.5%) of which are in the >4% frequency class. The genes that we have analyzed include 22 implicated in base excision repair, 12 disease genes associated with increased sensitivity to DNA damage, 14 genes related to DNA damage response genes from other species, 16 DNA polymerase subunits, 8 DNA replication checkpoint genes, 17 involved in homologous recombination, 11 mismatch excision repair genes,

18 mitotic spindle checkpoint genes, 11 genes involved in nonhomologous end joining, 32 nucleotide excision repair genes, 9 involved in post-replication repair, 42 genes with a suspected DNA repair function, and 90 genes in various other categories. Among the latter group are 39 candidate breast cancer susceptibility modifiers, which were included because BRCA1 and BRCA2 have roles in DNA repair and because in the absence of a fully assembled NF1 patient DNA panel, we practiced high throughput SNP genotyping using available somatic DNAs from approximately 450 early onset (diagnosis <40 years) breast cancer patients and about 400 ethnically matched controls (FitzGerald et al., 1997; and unpublished data). We obtained separate funding from the Avon Corporation to support this related project. Among the 793 missense and protein truncating SNPs identified thus far, 155 have a reported variant allele frequency >4%, 157 have an allele frequency between 1 and 4%, 200 are in the <1% allele frequency class, and for 281 SNPs the allele frequency remains unknown. We are most interested in SNPs in the >4% allele frequency category, since less common SNPs are unlikely to produce statistically significant results given the size of our patient panel.

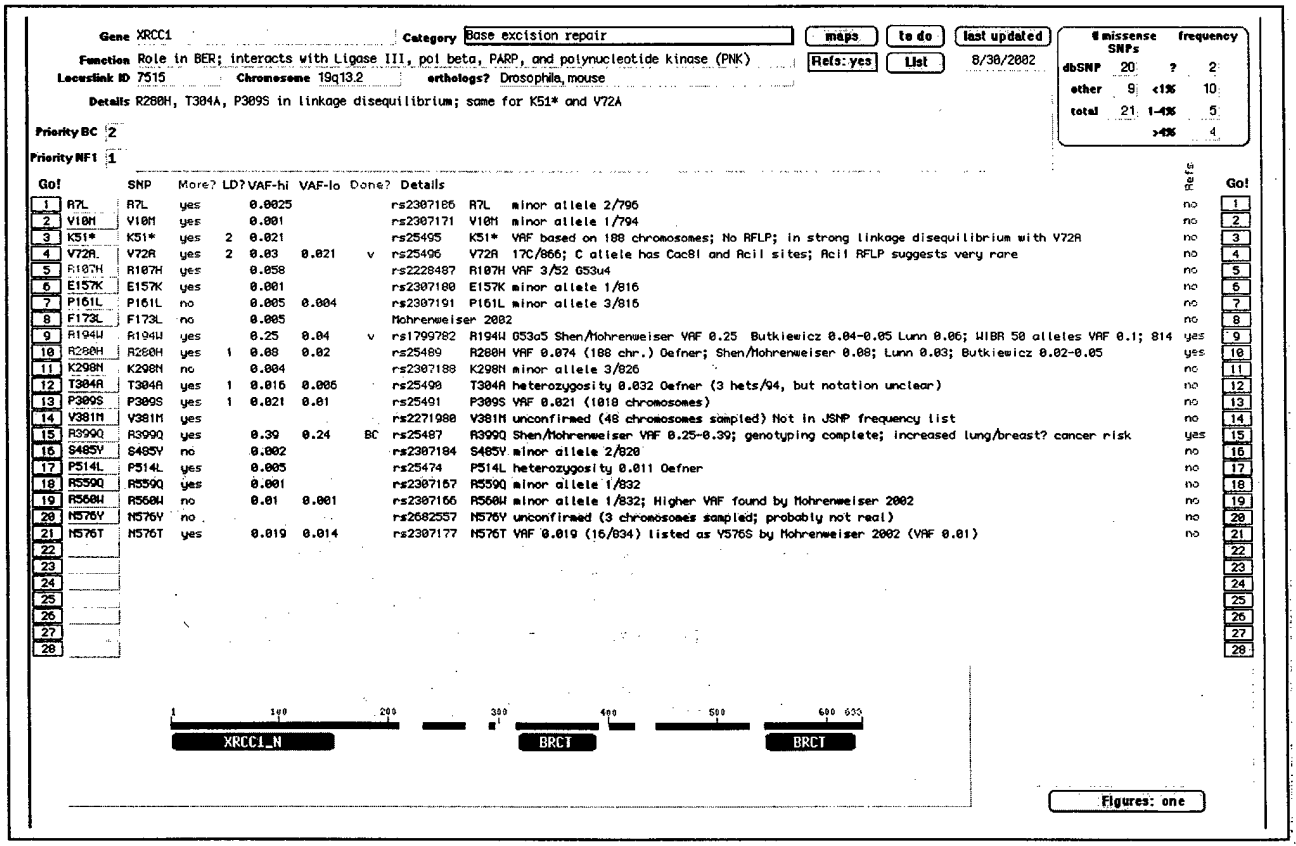

Figure 1. Main layout of SNP database for the *XRCC1* base excision repair gene. Relevant details are discussed in the text.

Although public databases such as dbSNP or GeneSNP continue to improve, data quality still leaves much to be desired (Marsh et al., 2002). Thus, a large proportion of database entries still represent SNPs identified exclusively *in silico*, often by comparing EST sequences. Typically, no population specific allele frequencies are known for such SNPs, the reality of which remains very much in question. Although this situation is slowly improving, online databases are still subject to frequent change, and in many cases long lists of SNPs are given without information as to whether they affect protein sequence. For all genes in our database we manually identified nonsynonymous SNPs. This involves making maps of the genes involved, which is a time consuming process. However, storing the maps used to identify SNPs as part of each gene's database record makes the evaluation of any future SNP straightforward. For typical SNPs, our database lists minor allele frequency, the sequence around the polymorphism, information on whether the SNP affects evolutionary conserved amino acids (determined by performing BLASTP searches; SNPs that alter evolutionary conserved amino acids will be analyzed with highest priority), details about genotyping methods (PCR primer design, etc), and abstracts of papers that cite the SNP (discovery, association and functional studies). The database also includes a computer generated domain structure of each protein (Figure 1), which helps to identify SNPs in potentially important protein segments. An important detail for database aficionados is that our overall database (current size 27.5 MB) consist of two integrated relational databases with gene-specific or SNP-specific information.

Anticipating the need to efficiently process and analyze bulk genotyping data, we also designed a third relational results database. This database centrally stores genotyping data and automatically calculates several basic statistical and genetic parameters from entered genotype data. Thus, entering observed genotypes calculates allele frequencies among cases and controls, expected allele frequencies based on Hardy-Weinberg equilibrium, $\chi 2$ *P* values for observed allele distributions assuming both recessive and dominant models, and odds ratios with 95% confidence intervals for all genotypes. Having a database that performs these basic calculations does not substitute for more sophisticated biostatistical analysis, but is invaluable in practice.

Beyond creating the required bioinformatics resources, much of the remainder of this project was contingent upon our ability to recruit 300 eligible NF1 patients within 15 months and up to 600 eligible patients within two years. Thus, **Task 2** involved the analysis of a limited number of MIT discovered missense SNPs in peripheral blood DNA samples from 150 high and 150 low neurofibroma number patients during months 1-15, while **Task 3** was to confirm any detected allele association in 300 additional high and low neurofibroma burden patients during the remaining nine months. **Task 4** was to perform protein truncation assays to detect additional loss-of-function mutations among genes that showed positive allele associations. Soon after the start of this project it became apparent that our recruitment goals, based on estimates provided by clinical collaborators, had been unrealistic. Thus, Dr. Korf at Boston Children's Hospital had estimated to contribute between 60 and 100 patients annually, and Dr. MacCollin at MGH had indicated she would contribute between 40 and 50 eligible patients each year. The remaining patients were to be recruited by advertising this study nationally.

At this time we have enrolled 76 patients by means of our original recruitment strategy. Thus, we enrolled 51 of 245 patients who contacted us in response to newsletter and other notices, 7 of 28 patients that were brought to our attention by Dr. Kurtz (whose eligibility criteria turned out to be significantly more relaxed than ours), 17 of 21 patients found eligible by Dr. MacCollin, and 1 of 2 of Dr. Korf's patients. Obviously, recruitment from all sources has run behind schedule. Among important reasons for this shortfall is that Dr. Korf gave up his directorship of the Boston Children's Hospital NF clinic just before the start of this project, and more recently has left Harvard Medical School for the University of Alabama at Birmingham. We also did not anticipate that the Army regulatory Compliance Department would not allow the recruitment of patients younger than 18 years of age, which excluded most patients seen at Boston Children's Hospital. Another problem was that Dr. MacCollin went without a clinical coordinator for nine month and has consistently recruited fewer patients that originally anticipated. Among the 245 patients who contacted us directly, more than half have so far received consent and blood drawing kits, but only 51 have returned consent forms and blood samples.

During the first year of this project it became obvious that our recruitment plan had two main deficiencies. First, we had anticipated recruiting about 50% of the required patients from just two local NF clinics, with the remaining 50% of patients to be recruited nationally in response to notices in patient organization news letters. The problems with this strategy were that it relied too heavily on the enthusiastic participation of just two clinics, and that outside patients were to be recruited based on self-reported neurofibroma numbers. This latter issue is a problem, since at least some patients desperately want to be part of any study that addresses a disease for which there is currently no cure. To circumvent both problems we attempted to enlist additional clinical collaborators, thus ensuring that eligibility would in most cases be evaluated by clinicians. Initially all domestic clinicians approached by us balked at participating in an Army funded study given the burdensome regulatory process. We had more success enlisting collaborators in Europe and Table 2, taken from a recently funded grant application, lists six clinicians who have agreed to recruit patients for this project. More recently, two domestic NF clinics have also agreed to recruit patients for this study, which is continuing with support from a newly funded Army Investigator Initiated Research Award (DAMD17-03-1-0438) to Andre Bernards.

Collaborator	Location	# DNAs available	# prospective patients
Evans, Gareth	Manchester, UK	0	150
Ferner, Rosalie	London, UK	0	>100
Lázaro, Conxi	Barcelona, Spain	55	30-60
Legius, Eric	Leuven, Belgium	0	>75
Mautner, Victor-Felix	Hamburg, Germany	288	300
Messiaen, Ludwine	Ghent, Belgium	50	50-70
Locally recruited	Boston, MA	66	100
Total		459	805-875

Table 2; Clinical collaborators and number of available or to-be-recruited eligible patients.

When initially contacted the six collaborators listed in Table 2 indicated they had DNAs from 393 eligible patients available for analysis. Beyond this number, they anticipated recruiting 705-770 more patients within three years. While these numbers argue that a 1200 patient panel (in order to have additional statistical power we have doubled the desired patient panel size) should be achievable without enlisting other collaborators, complications will inevitably arise. Thus, upon receiving Dr. Mautner's 288 samples it became clear that only about 50% of his patients were actually eligible or had sufficient (>5 microgram) DNA available. Dr. Messiaen will continue to

participate, but has recently moved from Belgium to the University of Alabama, Birmingham. The Army Regulatory Compliance Department has also ruled recently that we can no longer recruit domestic patients who contact us after learning about our study. Thus, in the future we can only recruit patients who have been seen at participating clinics. Finally, most collaborators have only agreed to participate if contributing patients anonymously circumvents the burdensome regulatory compliance paperwork. Thus, the experience gained during this pilot project has been invaluable and has prepared us to face the significant remaining patient recruitment challenges.

Our initial proposal was to genotype a limited number of missense SNPs discovered at MIT using a single base extension fluorescence resonance energy transfer (SBE-FRET) protocol. However, before the start of this project our collaborators at MIT had replaced SBE-FRET by a lower cost single base extension fluorescence polarization (SBE-FP) protocol. In this homogenous method SNP containing DNA segments are PCR amplified, followed by enzymatic degradation of primers and nucleotides, and extension of an unlabeled primer that abuts the SNP with fluorescent chain terminators. Incorporation of either one or both chain terminators is measured as an increase in fluorescence polarization (Kwok, 2002). In the first annual report we noted that our original plan to use MIT Genome Center equipment to read SNP genotypes turned out to be unworkable and that we had acquired our own MJ Tetrad thermal cycler and LJL-Analyst-AD 96/484 well fluorescence polarization plate reader. After spending considerable effort optimizing and evaluating the reliability of SBE-FP genotyping, we have reluctantly concluded that SBE-FP genotyping is not as problem-free as suggested. Thus, rather than close to 100% successful assays and >99% accuracy with little optimization (Hsu et al., 2001), only about 70% of our assays work even after extensive optimization, and average accuracy is no greater than 95%. We arrived at these numbers by genotyping multiple SNPs in parallel by SBE-FP and restriction fragment length polymorphism (RFLP) or allele specific PCR (ASPCR) methods. Using a combination of all three methods, we have successfully determined approximately 20,000 individual genotypes for 37 SNPs in 26 different genes in our breast cancer case/control panel. Results indicate that several missense SNPs in the Fanconi Anemia A (FANCA) gene that are in linkage disequilibrium and strongly associated with early onset breast cancer. For example, based on genotypes obtained for the FANCA Thr266Ala SNP in 551 patients and 393 controls, we calculated a Student t test p value for genotype distributions between cases and controls of 0.0002. Further suggesting a

significant association between FANCA missense alleles and early onset breast cancer, the odds ratio for the 266Thr allele was 1.75 (95% confidence interval 1.32-2.32). The breast cancer project was a direct spin-off of the work funded under this grant and these results are currently being prepared for publication.

Current SNP genotyping methods remain cumbersome and costly (typically $0.50 to $1,30 per genotype), making the analysis of candidate modifier SNPs the only practical approach. Although there is much excitement about matrix-assisted laser desorption-ionization time of flight (MALDI-TOF) mass spectroscopy based SNP genotyping, at $0.60 per four-fold multiplexed assay this method also remains too costly for anything but candidate SNP screens. Less biased genome-wide SNP association studies require knowledge of SNP haplotype block structure and of haplotype-tagging SNPs. This information is currently being generated by researchers at Perlegen and elsewhere. A genome-wide SNP association study also requires a method to reliable genotype large numbers (>10e5 per patient) of SNPs at low cost using limited amounts of DNA. No such method is yet available, although the Affymetrix GeneChip 10,000 SNP mapping array represents a significant step towards achieving this goal. Thus, we envisage that the patient DNA panel assembled during this project will eventually be used for comprehensive genome-wide SNP haplotype determinations.

Key Research Accomplishments

1. Designed and implemented patient information database
2. Designed and implemented Genome Stability Gene SNP database
3. Contacted >296 NF1 patients and enrolled 76 eligible individuals.
4. Identified clinical collaborators who will contribute >1000 additional patients
5. Determined >20,000 individual genotypes for 37 SNPs in 26 genes while evaluating genotyping methods.

Reportable Outcomes

- Meeting abstract and platform presentation. NNFF International Consortium for the molecular biology of NF1 and NF2. Aspen, CO. May 20-23, 2001.

- Meeting abstract and platform presentation. NNFF International Consortium for the molecular biology of NF1 and NF2. Aspen, CO. June 9-12, 2002.
- Meeting abstract and platform presentation. NNFF International Consortium for the molecular biology of NF1 and NF2. Aspen, CO. June 1-3, 2003.
- Patient database, Genome Stability Gene SNP database listing information on 793 missense SNPs in 293 candidate genome stability genes, and SNP Genotype Analysis database.
- Funded NIH R01 Grant Application. Title: Quantitative Phenotyping and Genotype-Phenotype Correlations in NF1; Principal Investigator: Bruce R. Korf. Results from the current project were used as preliminary data in this awarded application, which uses a discordant sib pair strategy to perform intrafamilial and interfamilial comparisons of dermal neurofibroma and café-au-lait macule numbers for identification of modifier loci.
- NIH R01 Grant Application. Title: Studies of neurofibromatosis-1 modifier genes. Principal Investigator: Andre Bernards. Results from the current project were used as preliminary data in this application, whose main aims include allele association studies to evaluate three classes of potential neurofibroma burden modifiers. This proposal received a 35% priority score upon first review.
- Army NF Research Program Investigator-Initiated Research Proposal. Title: Studies of neurofibromatosis-1 modifier genes. Principal Investigator: Andre Bernards. Results from the current project were used as preliminary data in this recently funded application (DAMD17-03-1-0438), which has complete scientific and budgetary overlap with the NIH R01 application listed above.

Conclusions

The main goals of this 2 year project were to collect somatic DNAs from 600 NF1 patients that represent the top and bottom 20% of neurofibroma burden and to use this resource to evaluate whether protein-altering alleles of genes implicated in maintaining genome stability are associated with a high or low neurofibroma burden. We encountered several major problems during the execution of this project. Firstly, our plan to genotype missense SNPs in candidate modifier genes identified in a SNP discovery screen at MIT ran into problems when it became apparent that only a small fraction of candidate modifier genes had been analyzed in the MIT screen. This required us

to perform time consuming data mining in order to identify a comprehensive set of candidate modifier alleles. Secondly, our plan to use MIT Genome Center equipment to read SNP genotypes turned out to be impractical, requiring is to buy our own Analyst-AD 96/384 well fluorescence polarization plate reader. Thirdly, notwithstanding published reports to the contrary, in our hands SBE-FP genotyping is not robust and requires too much optimization to allow efficient analysis of multiple SNPs. Although we remain interested in evaluating other methods, we have currently settled on labor intensive but reliable RFLP or ASPCR genotyping methods to achieve our goals. Finally, our patient recruitment plan, designed with significant input from the Chair of the Medical Affairs Committee of the National Neurofibromatosis Foundation, turned out to be inadequate. While these factors made it impossible to reach our stated goals, the experience gained during this pilot project has been invaluable and has allowed us to obtain additional grant support to continue our efforts to identify genetic modifiers of neurofibroma burden in NF1 patients.

References

Easton, D. F., Ponder, M. A., Huson, S. M., and Ponder, B. A. (1993). An analysis of variation in expression of neurofibromatosis (NF) type 1 (NF1): evidence for modifying genes. Am J Hum Genet *53*, 305-313.

FitzGerald, M. G., Bean, J. M., Hegde, S. R., Unsal, H., MacDonald, D. J., Harkin, D. P., Finkelstein, D. M., Isselbacher, K. J., and Haber, D. A. (1997). Heterozygous ATM mutations do not contribute to early onset of breast cancer. Nat Genet *15*, 307-310.

Hsu, T. M., Chen, X., Duan, S., Miller, R. D., and Kwok, P. Y. (2001). Universal SNP genotyping assay with fluorescence polarization detection. Biotechniques *31*, 560, 562, 564-568

Huson, S. M., and Hughes, R. A. C., eds. (1994). The neurofibromatoses: A pathogenetic and clinical overview, First edn (London, Chapman & Hall Medical).

Kwok, P. Y. (2002). SNP genotyping with fluorescence polarization detection. Hum Mutat *19*, 315-323.

Marsh, S., Kwok, P., and McLeod, H. L. (2002). SNP databases and pharmacogenetics: A great start but a long way to go. Hum Mutation *20*, 174-179.

Szudek, J., Birch, P., Riccardi, V. M., Evans, D. G., and Friedman, J. M. (2000). Associations of clinical features in neurofibromatosis 1 (NF1). Genet Epidemiol *19*, 429-439.

Appendices

None