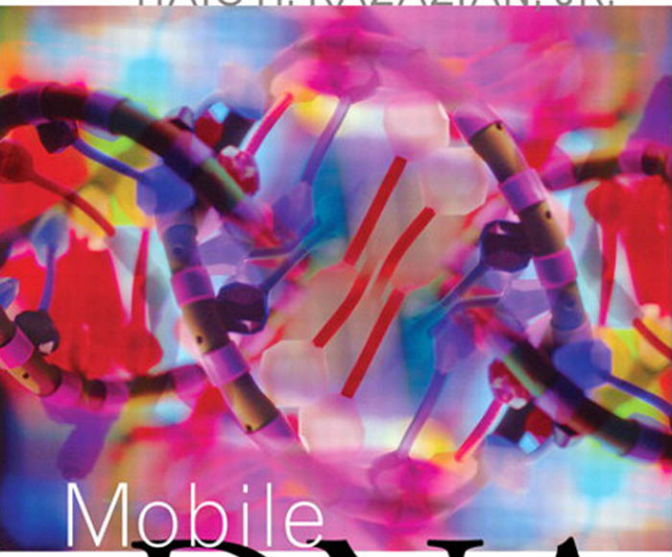
HAIG H. KAZAZIAN, JR.



Mobile DNA

Finding Treasure in Junk

Mobile DNA

This page intentionally left blank

Mobile DNA

Finding Treasure in Junk

Haig H. Kazazian, Jr.

Vice President, Publisher: Tim Moore
Associate Publisher and Director of Marketing: Amy Neidlinger
Acquisitions Editor: Kirk Jensen
Editorial Assistant: Pamela Boland
Operations Manager: Gina Kanouse
Senior Marketing Manager: Julie Phifer
Publicity Manager: Laura Czaja
Assistant Marketing Manager: Megan Colvin
Cover Designer: Gary Adair
Managing Editor: Kristy Hart
Senior Project Editor: Lori Lyons
Copy Editor: Language Logistics, LLC
Proofreader: Sheri Cain
Indexer: Angela Martin
Compositor: Nonie Ratcliff
Manufacturing Buyer: Dan Uhrig

© 2011 by Pearson Education, Inc.
Publishing as FT Press
Upper Saddle River, New Jersey 07458

FT Press offers excellent discounts on this book when ordered in quantity for bulk purchases or special sales. For more information, please contact U.S. Corporate and Government Sales, 1-800-382-3419, corpsales@pearsontechgroup.com. For sales outside the U.S., please contact International Sales at international@pearson.com.

Company and product names mentioned herein are the trademarks or registered trademarks of their respective owners.

All rights reserved. No part of this book may be reproduced, in any form or by any means, without permission in writing from the publisher.

Printed in the United States of America

First Printing April 2011

ISBN-10: 0-13-707062-4
ISBN-13: 978-0-13-707062-6

Pearson Education LTD.
Pearson Education Australia PTY, Limited.
Pearson Education Singapore, Pte. Ltd.
Pearson Education Asia, Ltd.
Pearson Education Canada, Ltd.
Pearson Educación de Mexico, S.A. de C.V.
Pearson Education—Japan
Pearson Education Malaysia, Pte. Ltd.

Library of Congress Cataloging-in-Publication Data

Kazazian, Haig H. (Haig Hagop), Jr., 1937- author.
Mobile DNA : finding treasure in junk / Haig Kazazian.
p. ; cm.
Includes bibliographical references.
ISBN-13: 978-0-13-707062-6 (hardback : alk. paper)
ISBN-10: 0-13-707062-4 (hardback : alk. paper)
1. Insertion elements, DNA. 2. Transposons. 3. Recombinant, DNA. 4. Genomics. I. Title.
[DNLM: 1. DNA Transposable Elements. 2. DNA, Recombinant—physiology. 3. Genomics—methods. 4. Sequence Analysis, DNA. QU 470]
QH462.I48K39 2011
572.8'69—dc22

2010050899

To Lilli, my loving wife of so many years,
who put up with my idiosyncrasies
and encouraged me to write this book.

This page intentionally left blank

Contents

Acknowledgments

I thank my early mentors, Lafayette Noda and Lucille Smith, at Dartmouth; Barton Childs at Johns Hopkins; and Harvey Itano at the NIH. I thank colleagues who helped me in entering the mobile DNA field, especially Maxine Singer, Allan Scott, and Jef Boeke. I thank all the lab members, both those mentioned and those unmentioned, for their tireless efforts to contribute to the mobile DNA field. I thank other colleagues, John Moran, Nancy Craig, Vivian Cheung, and Eric Ostertag, for helpful comments and critiques of the book. I also thank Adam Ewing, Dustin Hancks, and Lauren Zeitel for their advice and comments.

About the Author

Haig Kazazian was born and raised in Toledo, Ohio. After attending public schools there, he received his A.B. degree from Dartmouth College in 1959. He then attended Dartmouth Medical School, a two-year school at the time, and finished his M.D. degree at Johns Hopkins University School of Medicine. At Hopkins, he met his wife of nearly 50 years and married during his internship in Pediatrics at the University of Minnesota Hospital. After two years training in Minneapolis, he returned to Johns Hopkins for a two-year fellowship in genetics with Barton Childs, M.D. He then trained for two years in molecular biology in the lab of Harvey Itano, M.D., at the NIH. After a third year of Pediatric training at Johns Hopkins, he joined the faculty there in 1969. He rose through the ranks to become a full professor in 1977, and at that time, he headed the Pediatric Genetics Unit. In 1988, he became Director of the Center for Medical Genetics at Johns Hopkins.

After 25 years on the Hopkins faculty, he was recruited to the University of Pennsylvania School of Medicine as Chair of the Department of Genetics in 1994. At Penn, he recruited 10 young faculty to the department. In 2006, he stepped down as department chair, but remained as the Seymour Gray Professor of Molecular Medicine in Genetics until 2010. In July 2010, he returned to Johns Hopkins as a Professor in the Institute of Genetic Medicine.

Dr. Kazazian is still heavily involved in molecular genetic research, concentrating for the past 20 years on mammalian and human transposable elements, or "jumping genes." Prior to 1988, he characterized much of the variation in the cluster of genes involved in production of the beta chain of human hemoglobin. With Stuart Orkin at Harvard, his work led to the nearly complete characterization of the mutations causing the β-thalassemias, common anemias in regions of the world endemic for malaria.

Dr. Kazazian is a member of a number of national organizations, including the Institute of Medicine of the National Academy of Sciences and the American Academy of Arts and Sciences. He has received a number of honors for his research, most notably the 2008 William Allan Award, the top honor of the American Society of Human Genetics.

Preface
Thoughts on doing science

Before diving into the subject of mobile DNA and my adventures in the field, I'd like to provide a few personal tips from my experience on working with some success in science for the past 45 years. Doing science is often very difficult and extremely hard work, requiring long hours. In my view, the first thing necessary is very high personal motivation. My original and long-time mentor, Barton Childs, an esteemed Professor of Pediatrics and the "father of pediatric genetics," always used to say, "You've got to burn to do research!" You can't go at it with a half-hearted enthusiasm or self-doubt.

If you do have high personal motivation, you then need to get excellent training, both in didactic class work and in the nuts and bolts of how to do research. You need to find a subject or area that really interests you, no matter what the field. Then find the investigator who is doing excellent research in your field of interest, hopefully someone at the forefront, but also consider that the person's lab is not too large so that he or she will have sufficient time to spend with you. You want someone who will discuss your research with you on a daily basis, perhaps so much so that you feel that he or she is pestering you all the time for new data. That kind of attention means that that individual has great interest in your work. You need that kind of person for both your predoc and your postdoc training. Your trainers also need to be available for discussion of all kinds of problems, both those that you face in the lab and those that are related to other aspects of your life.

Next, you need a dependable mentor. Your mentor could be either your predoc or your postdoc trainer, or it could be a member of your thesis committee or another senior investigator from down the hall. However, you'd like a mentor that you can carry over from your

training days into your first 5–10 years as a faculty member. That mentor can help you with all kinds of problems and questions, giving advice for how to approach various professional and daily life situations. Having an interested, accessible, and experienced mentor is crucial to success in science. Behind every good scientist is an outstanding mentor. I certainly had one in Barton Childs, even if I didn't follow his advice at every turn. He was probably my major mentor for at least 20 years.

I've talked about mentorship from the aspect of the trainee, but what about the importance of being a good mentor? From your first academic job to becoming a long-term lab director, you have the responsibility for mentoring predoc and postdoc trainees. I have usually found it rewarding to give trainees considerable independence, letting them pick their own problems from a wide variety of problems available in the lab. This works well when the trainee is very bright and picks one or more problems that are of real interest to the lab director. When the problem is of little interest to the lab director, there is a good chance that the work will flounder. However, if the problem is important to the mentor, the mentor will add ideas and enthusiasm to the work. I have dealt with both situations over my career, as the reader will discover in this book. A third situation occurs when the student or postdoc needs to finish a period of training and hasn't had much success up to that point. The trick at that time is to find a project that is important for the field (so that the student will take pride in his or her accomplishment), has a clear-cut endpoint, and uses techniques already available in the lab. The design of this project usually requires considerable input from the lab director.

Then there is the question of how one should approach other scientists. Should one be open in discussing new data even with colleagues in the same field, or should one be secretive to avoid being "scooped?" My view has always been that it is better to be open but prudent. It is good to discuss your unpublished work at meetings. If your work is important, your colleagues will respect you for talking

about new data and not rehashing work that has been published and that they've heard previously. Moreover, it is very, very rare that another investigator can start a new experimental tack or line and actually beat you to publication. After all, you've probably been working on that same question for a year or two, so you've got a major head start. It's a rare investigator indeed who would sail off in a new direction hoping to beat you to publication of data that you've just presented. As a general rule, openness in presenting and discussing new data is the best approach.

A corollary to openness is to discuss your science with a wide range of other scientists, including those within your immediate field, e.g., mobile DNA, those in the broader field, e.g. human genetics, and those in other fields of biomedical science, e.g., immunology or developmental biology. You never know from where the next good experimental idea will come. The reader will find throughout this book that members of my lab and I personally have gotten ideas from a wide variety of sources who are mentioned at some future time. This plethora of good ideas has come from discussing the work with a large number of other scientists and sources and being as open as possible to new ideas.

I once knew a well-trained, smart young researcher who had a great deal of trouble gaining traction in his field. I always thought that his problem was that he stayed in his lab and did not seek discussion of his science with colleagues. At the other extreme was and still is the Medical Research Council (MRC) laboratory at Cambridge, England, whose investigators have had enormous success over many decades. The Cambridge MRC labs have housed a number of Nobel Laureates, including Francis Crick, Fred Sanger, Sydney Brenner, Aaron Klug, and others. After spending a few months at the MRC early in my career, I felt that a major factor in the success of that lab was the English tradition at that time of a common coffee break in the morning and a common tea break in the afternoon. At 10:30 AM, every investigator, from the trainees to the most senior people, would

gather in the cafeteria for morning coffee and, importantly, discuss science for 30 minutes over coffee. A similar gathering would occur at 3:30 PM over afternoon tea. The number of great new ideas passed from one investigator to another, from past and future Nobel Prize winners to beginning postdocs, and vice versa, was astonishing. Open discussion of science is wonderful for the development of new ideas.

Now I'd like to make a general comment on picking problems in your field on which to work. I've always believed that the problem should be important but potentially solvable with hard effort. All researchers are gamblers. A colleague used to tell me to pick problems with 5 to 1 or 10 to 1 odds of success. Those problems were about right in terms of difficulty. Odds of 50 to 1 or 100 to 1 were too long, and success on those problems was too unlikely. Odds of 2 to 1 meant that the problem was too easy and relatively unimportant, so called "low-hanging fruit." I've also felt that it is best to pick problems that are logical next steps in the project but are important to the field and have those reasonable odds of success, which would be 5–10:1.

My last point is to keep one's mind alert for possible collaboration. Collaborations with other scientists should be welcomed as a way to broaden one's scientific outlook and scope. If two investigators have differing expertise that can be applied to solve a particular problem, this is an ideal situation for collaboration. I once heard it said that collaboration finds its own level, meaning that in order to work best, collaborators should be on the same level of experience and respect in the field. In this way, I've had good collaborations as a postdoc with another postdoc and as a senior scientist with other senior scientists. Many of these collaborations are discussed throughout the book.

This page intentionally left blank

1

Introduction to Mobile DNA

Charles Darwin would be surprised. Indeed, even present day scientists are surprised by the existence of mobile DNA. Consider the skepticism within the scientific community that greeted Barbara McClintock, already a highly-respected scientist, when she announced that she had found what appeared to be mobile DNA in maize plants (McClintock, 1950). DNA was the genetic material, so it must be static, stable, and immobile. The mutation rate had been determined to be ~10^{-8} per nucleotide, or building block, of DNA per generation—very low indeed. How and why would some DNA move from place to place in a genome? Scientists are still grappling with these questions. Two hundred years removed from Darwin's birth, and we're still wondering how mobile DNA with all its detrimental effects on organisms could have reached such high proportions in the genomes of mammals and plants. Yet mobile DNA is found in all forms of living things, including plants, animals, bacteria, and archea. The genome seems to cherish its ability to make rapid changes by rearranging some of its parts as opposed to the slow change afforded by the nucleotide mutation rate.

One theme of this book is that biological scientists have come to expect the unexpected. The study of living things is full of surprises. One of them is the prevalence of mobile DNA in genomes. Another is that most genes are broken up by sections of DNA called introns that need to be removed at the RNA stage in order for the genes to function. A third is that the protein-coding regions of genes make up a very small fraction of mammalian genomes. A fourth surprise is the importance of reverse transcriptase, the enzyme that synthesizes DNA from an RNA template. These are just a few examples of old surprises, or unexpected findings, that have now become hard facts in

all biology textbooks. Many more will be highlighted in the research adventures outlined in this book. These "unexpected observations" provide excitement and anticipation for even the most experienced researchers. What finding will be the next to shatter our present view of the biological world? One can be sure that the future will bring many more surprises to delight the graduate student just beginning his or her studies.

Prior to 1970, scientists thought that the genome was composed mostly of genes lined up like balls on a string with some repetitive DNA in between the balls. Then in the late 1970s, introns were found to break up the regions of genes that encode proteins (Berget et al., 1977; Chow et al., 1977). Protein-coding regions were disrupted by intervening sequences (introns) that required removal from pre-messenger RNA before the intact protein could be synthesized. Soon, we knew that introns were much larger than protein-coding regions, then called exons. The DNA between the genes along with most of the intronic sequences of genes was thought to be functionless, and was called "junk DNA" (Orgel and Crick, 1980). However, now we know that introns make up about 30% of human and mammalian genomes, and exons encode only between 1 and 2% of the human genome (Lander et al., 2001). What a comedown for protein-coding regions! Thus, over 98% of human DNA had been dismissed as "junk."

Transposable elements were then found in human DNA, and this active mobile DNA along with the remnants of many transposition events over hundreds of millions of years is now known to account for at least 50% of human genomic DNA. This transposable element DNA, both those relatively few sequences that are presently mobile, and the many remnants of old events are now demonstrating function. However, this function is evident only in the many ways mobile DNA can modify the genome over evolutionary time. It can be co-opted for useful purposes but has not yet been definitively shown to have a useful function in the individual organism. Moreover, DNA encoding small RNAs of different types and functions has been discovered amidst the "junk." Enhancer sequences at great distances from the genes upon which they act are being found continually. Segmental duplications of hundreds to many thousands of nucleotide pairs of DNA are strewn around the genome and are further grist in

the mill of genome plasticity and malleability. The bottom line is that "junk" DNA is gradually being eroded away as function is found for a greater and greater fraction of genomic DNA. In this book, I concentrate on the "junk" DNA that is mobile or has been over the millennia. This is the DNA that those of us in the mobile DNA field have come to treasure.

In the next several chapters, I provide details on important topics in the mobile DNA field as well as discuss a number of top scientists who have been pioneers in many areas involving mobile DNA. I then discuss the state of the human mobile DNA field prior to my involvement in it, what led to my fascination with mobile DNA, and why I jumped at the chance to work on it when the opportunity presented itself. Later, I discuss many of the people who worked in my lab up to the present time, their most important work, and the relationship of that work to what is known about L1 biology today. This is followed by important findings of other labs working on mammalian mobile DNA, ending with some thoughts about the future of the field. Yes, DNA as genetic material would have surprised Charles Darwin, but mobile DNA would have really made his head spin!

This page intentionally left blank

Index

C

D

E

H

I

J–K

L

M

N

O

P

Q–R

S

T

U–V

W

X–Y–Z