

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

FORESIGHT DIAGNOSTICS INC.
Petitioner

vs.

PERSONALIS, INC.
Patent Owner

DECLARATION OF INGRID HSIEH-YEE, PH.D.

TABLE OF CONTENTS

I.	Introduction	3
A.	Qualifications and Professional Experience	3
B.	Scope of This Declaration.....	5
C.	Evidence Considered in Forming My Opinions	6
II.	Authenticity and Public Availability of <i>Leary</i> (Ex.1002).....	7
A.	Authentication.....	7
B.	PubMed Record for <i>Leary</i>	9
C.	Internet Archive Access	12
D.	Usage Record.....	14
E.	Summary of My Opinion on <i>Leary</i>	14
III.	Authenticity and Public Availability of <i>Chan</i> (Ex.1008).....	15
A.	Authentication.....	15
B.	PubMed Record for <i>Chan</i>	16
C.	Swanton Commentary on <i>Chan</i>	20
D.	Copyright registration record.....	23
E.	Summary of My Opinion on <i>Chan</i>	24
IV.	Conclusion.....	24

I. Introduction

I, Ingrid Hsieh-Yee, Ph.D., do hereby declare as follows:

1. I have personal knowledge of all of the matters about which I discuss in this declaration.

2. I have been retained as an independent expert witness on behalf of Foresight Diagnostics, Inc. (“FORESIGHT”) for proceedings before the United States Patent and Trademark Office (“PTO”) regarding the authenticity and public availability of certain prior art references in relation to US Patent No. 11,384,394.

3. I am being compensated for my work in this matter at my customary hourly rate. I am also being reimbursed for any reasonable expenses associated with my work and testimony in this investigation. My compensation is not contingent on the results of my study, the substance of my opinions, or the outcome of this matter.

A. Qualifications and Professional Experience

4. My complete qualifications and professional experience are described in my academic curriculum vitae (**Appendix A**). The following is a brief summary of my relevant qualifications and professional experience.

5. I was a professor in the Department of Library and Information Science at the Catholic University of America for 32 years and retired at the end of August 2022. I have experience working in an academic library, a medical library, and a legislative library. I hold a Ph.D. in Library and Information Studies from the

University of Wisconsin-Madison and a Masters in Library and Information Studies from the University of Wisconsin-Madison.

6. I am an expert on library cataloging and classification and have published two editions on this subject, *Organizing Audiovisual and Electronic Resources for Access: A Cataloging Guide* (2000 and 2006). I taught a variety of courses, including Cataloging and Classification, Advanced Cataloging and Classification, Organization of Internet Resources, Organization of Information, Digital Content Creation and Management, Internet Searches and Web Design, Information Literacy Instruction, Advanced Information Retrieval and Analysis Strategies, and The Information Professions in Society. I am familiar with metadata schema design and implementation. In my teaching I covered the design and implementation of metadata in databases, search engines, digital repositories, digital libraries, and digital archives. I also covered database design and how information organization affects the discovery and access to digital resources on the Internet. My research interests cover cataloging and classification, information organization, metadata, information retrieval, information architecture, digital collections, scholarly communication, user interaction with information systems, and others.

7. I am personally familiar with many online catalogs, databases, search engines, digital libraries and digital archives. In preparing for this declaration, I used the following authoritative information systems to search for records:

- Google Scholar (<https://scholar.google.com/>)
- Internet Archive (<https://archive.org/>)
- Public catalog of the U.S. Copyright Office
(<https://cocatalog.loc.gov/>)
- PubMed (<https://pubmed.ncbi.nlm.nih.gov/>)
- PubMed Central (<https://www.ncbi.nlm.nih.gov/pmc/>)

B. Scope of This Declaration

8. I have been asked to offer an opinion on the authenticity and public availability date of the following documents:

- Leary, R. J., Kinde, I., Diehl, F., Schmidt, K., Clouser, C., Duncan, C., ... & Velculescu, V. E., (2010, Feb. 24), Development of personalized tumor biomarkers using massively parallel sequencing, *Science Translational Medicine*, vol. 2, no. 20, page 20ra14, **Ex.1002** (“Leary”);
- Chan, K. A., Jiang, P., Zheng, Y. W., Liao, G. J., Sun, H., Wong, J., ... & Lo, Y. M. D., (2013, January 1), Cancer genome scanning in plasma: detection of tumor-associated copy number aberrations, single-nucleotide variants, and tumoral heterogeneity by massively parallel sequencing, *Clinical Chemistry*, vol. 59, no. 1, pp. 211-224, **Ex.1008** (“Chan”).

C. Evidence Considered in Forming My Opinions

9. In the preparation of this declaration, I have reviewed the documents referenced below and any other documents I reference herein, and each of these is a type of material that experts in my field would reasonably rely upon when forming their opinions:

- (1) The documents referenced above in Section I.B;
- (2) PubMed metadata record for *Leary*, available at <https://pubmed.ncbi.nlm.nih.gov/20371490/>, accessed and obtained on October 17, 2022, **Appendix 1002A**;
- (3) Website of *Science Translational Medicine*, archived by the Internet Archive on February 25, 2010, accessed and obtained on October 17, 2022, **Appendix 1002B**;
- (4) Citations to *Leary*, obtained from Google Scholar, **Appendix 1002C**;
- (5) PubMed metadata record for *Chan*, available at <https://pubmed.ncbi.nlm.nih.gov/23065472/>, accessed and obtained on October 10, 2022, **Appendix 1008A**;
- (6) Swanton, C., (2013), Plasma-derived tumor DNA analysis at whole-genome resolution, *Clinical Chemistry*, vol. 59, no. 1, pp. 6-8

(previously published online on October 31, 2012), available at the publisher website at

<https://academic.oup.com/clinchem/article/59/1/6/5622134>,

accessed and obtained on October 10, 2022, **Appendix 1008B**;

- (7) PubMed metadata record for *Swanton*, available at <https://pubmed.ncbi.nlm.nih.gov/23115052/>, accessed and obtained on October 10, 2022, **Appendix 1008C**;

- (8) Copyright registration record for the print issue of vol. 59, no. 1 of *Clinical Chemistry*, available at <https://cocatalog.loc.gov/cgi-bin/Pwebrecon.cgi?v1=92&ti=76,92&Search%5FArg=clinical%20chemistry&Search%5FCode=TALL&CNT=25&PID=RnVcF4wzREI3T2XU7XOyP4XtD9Rk&SEQ=20221010200253&SID=1> from the public catalog of the U.S. Copyright Office, accessed and obtained on October 10, 2022, **Appendix 1008D**.

II. Authenticity and Public Availability of *Leary* (Ex.1002)

A. Authentication

10. Ex.1002 is a true and correct copy of Development of personalized tumor biomarkers using massively parallel sequencing, (“*Leary*”), *Science Translational Medicine*, vol. 2, no. 20, pp. 20ra14-20ra14, that I obtained from the publisher. I began my research on Google Scholar by searching by the title of *Leary*

for records, and a retrieved record led me to the publisher’s website for *Leary* at <https://www.science.org/doi/abs/10.1126/scitranslmed.3000702>. The “check access” link on this website brought up a page that informed users that the article was free to AAAS members, and “[a]s a service to the community, this article is available for free. Login or register for free to read this article.” As a member of AAAS I then logged into the system to download the full text of *Leary*. This AAAS copy is presented as **Ex.1002** in this declaration.

11. **Ex.1002** is a true and correct copy of *Leary* obtained from the publisher AAAS. Page 1 shows the title, authors, abstract and a note showing *Leary* was published on “24 February 2010; Volume 2 Issue 20 20ra14” and is available on www.ScienceTranslationalMedicineorg. Page 6 shows supplementary materials of this article include three tables and two figures that are available online. Page 7 shows the manuscript was submitted on “7 December 2009, accepted 4 February 2010, published 24 February 2010” with “10.1126/scitranslmed.3000702” as its digital object identifier. **Ex.1002** shows *Leary* has a total of seven pages, two tables, four figures and 30 references. The AAAS website for *Leary* includes a download link for the supplementary materials. I downloaded these tables and figures and included them as pages 9 to 15 of **Ex.1002**.

12. The publisher’s copy of *Leary* (**Ex.1002**) is in a condition that creates no suspicion about its authenticity. Specifically, this copy is not missing any

intermediate pages, the text on each page appears to flow seamlessly from one page to the next, and there are no visible alterations to the document. Moreover, **Ex.1002** is in the custody of the publisher, a place where, if authentic, it would likely be found. Accordingly, I see no reason to question the authenticity of this copy of *Leary* (**Ex.1002**).

B. PubMed Record for *Leary*

13. To determine whether *Leary* was indexed, I searched PubMed by the article title for records and located a record at <https://pubmed.ncbi.nlm.nih.gov/20371490/>. This PubMed record is presented as **Appendix 1002A** in this declaration. PubMed is an authoritative system that contains “more than 34 million citations for biomedical literature from MEDLINE, life science journals, and online books” and PubMed records may include links to “full text content from PubMed Central and publisher web sites” (source: <https://pubmed.ncbi.nlm.nih.gov>). I personally located, identified and obtained the PubMed metadata record for *Leary* (**Appendix 1002A**). I saved the metadata record in PDF and also saved the record in PubMed format to show the record details in data fields.

14. **Appendix 1002A** is a true and correct copy of the PubMed metadata record for *Leary*. The top of page 1 shows a full text link to “Science Translational Medicine AAAS” that takes users to the publisher’s website for *Leary*. Page 1 also

shows the title, authors, and an abstract. The area above the title shows a note of “Sci Transl Med. 2010 Feb 24; 2(20):20ra14. Doi: 10.1126/scitranslmed.3000702” that indicates *Leary* was published on February 24, 2010, in vol. 2, no. 20, pages 20ra14 of *Science Translational Medicine*, with a doi (digital object identifier) of “10.1126/scitranslmed.3000702.” The authors, title, journal title, issue numbering, page numbers and doi match the information presented in the publisher’s copy of *Leary* (**Ex.1002**), confirming this record represents *Leary*. Page 1 also shows the PMID (PubMed Identifier) of this metadata record is “20371490” and a copy of *Leary* is freely available at PMC, the PubMed Central digital archive. The PMC record’s identifier is 2858564.

15. I obtained PMC2858564 from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2858564/> and included a screen capture of the top this record as page 3 of **Appendix 1002A**. This record shows the PMC copy is an author manuscript that became available for public access in PMC on “2010 Aug 24.” The archived copy is available as a webpage, a PDF, and a PubReader file.

16. Pages 4 to 5 of **Appendix 1002A** are the PubMed record in PubMed format to show data in specific fields. I use the “MEDLINE/PubMed Data Element (Field) Descriptions” (<https://www.nlm.nih.gov/bsd/mms/medlineelements.html>) to interpret the record. The “PMID” field shows “20371490” as the PubMed identifier

for the record for *Leary*. The “STAT” field shows the status of the record is “MEDLINE” (meaning Medical subject headings were assigned to the article), and the DCOM field shows this index record was completed on “20100615” (*i.e.*, June 15, 2010). The first IS field shows the ISSN (International Standard Serial Number) for the electronic version of this journal is “1946-6242” and the ISSN for the print journal is “1946-6234.” Fields VI (volume), IP (issue), DP (date of publication), Ti (title), PG (pages) and JT (journal title) show *Leary* was published in vol. 2, no. 20 (2010 Feb 24) of *Science Translational Medicine* as page 20ra14. The first FAU (full author) field shows “Leary, Rebecca J” as the first author. Field CRDT (creation date) and Field EDAT (entrez date) show the metadata record was added to PubMed on “2010/04/08” (*i.e.*, April 8, 2010), and Field MHDT (MeSH date) shows medical subject headings were added to the record on “2010/06/16” (*i.e.*, June 16, 2010) to elevate this record into MEDLINE (source: <https://pubmed.ncbi.nlm.nih.gov/help/>). The PubMed record shows 11 medical subject headings were assigned to indicate the subjects of the *Leary* article. Field “AID” (Article identifier) shows the doi of *Leary*.

17. This PubMed metadata record (**Appendix 1002A**) made *Leary* discoverable in PubMed on April 8, 2010, when it was entered into PubMed. Interested users would have been able to search for *Leary* in PubMed by the authors, title, journal title, ISSN, and abstract. After medical subject headings were added to

enrich the PubMed record on June 16, 2010, interested users would have been able to conduct subject searches to discover *Leary*.

C. Internet Archive Access

18. To determine the availability of *Leary* on the Internet I searched Internet Archive for records. Internet Archive is a digital library that provides access to websites, webpages, and other web resources, and provides free access to its contents for researchers, scholars, and the general public. **Appendix 1002B** is a true and correct copy of the website of *Science Translational Medicine* archived by the Internet Archive on February 25, 2010. I personally located, identified and obtained this record. This is the type of record experts in my field would reasonably rely upon when forming their opinions.

19. Page 1 of **Appendix 1002B** is a Wayback Machine calendar that shows the original URL of this website was <http://stm.sciencemag.org/> and the website was captured 2,318 times from “13 Oct 2009” to “8 Oct 2022.” The calendar shows this archived copy was captured on “Feb 25 2010” and is available at web.archive.org/web/20100225013602/http://stm.sciencemag.org/. I have used Internet Archive in the past and am familiar with the Internet Archives’ practice of assigning an extended URL to an archived file in the format of [http://www.archive.org/web/\[Year in yyyy\]\[Month in mm\]\[Day in dd\]\[Time code in hh:mm:ss\]/\[Archived URL\]](http://www.archive.org/web/[Year in yyyy][Month in mm][Day in dd][Time code in hh:mm:ss]/[Archived URL]) to show the point of time an archived file was

captured. The URL of the archived website means this website was archived on “20100225013602” (*i.e.*, February 25, 2010, 1 a.m., 36 minutes and 2 seconds). Because Internet Archive does not capture web resources as soon as they appear on the Internet, the archive date means the archived resource would have been available for public access no later than the archive date. The archive date of **Appendix 1002B** therefore means that this website would have been available for public access no later than February 25, 2010.

20. Appendix 1002B shows “Development of Personalized Tumor Biomarkers Using Massively Parallel Sequencing” was a research article highlighted on the website by February 25, 2010. Pages 2 and 3 show the website saved in PDF and page 3 shows this archived website has a “2010” copyright date, with the American Association for the Advancement of Science as the copyright holder. The link of this research article retrieves an archived copy of the webpage for this article. Page 4 of **Appendix 1002B** is a screen capture of this webpage with a Wayback Machine calendar that shows this archived copy was captured on “Jun 09 2012.” Pages 5 and 6 are the archived webpage saved in PDF. The title, authors, abstract and publication history on this page matches the information presented in the AAAS copy of Leary (**Ex.1002**). The “Read the Full Text” link on page 5 leads to the webpage for the full article (page 7 of **Appendix 1002B**) that informs visitors that they need to “Register for Free or Subscribe/Join AAAS to View Full Text.” These

archived records from the Internet Archive inform my opinion that *Leary* was available on the Internet no later than February 25, 2010, to registered users and AAAS members.

D. Usage Record

21. Actual usage of an article is reflected by other articles that makes reference to it. The citation history of Google Scholar shows *Leary* has been cited 629 times, including 24 times in 2010. **Appendix 1002C** presents seven citations published in 2010 to demonstrate early usage. The earliest citation appeared in March 2010. Page 2 is a screen capture of Google Scholar showing 629 citations made reference to *Leary*, and page 3 is a screen capture showing 24 citations appeared in 2010.

E. Summary of My Opinion on *Leary*

22. Taken together, the AAAS copy of *Leary* (**Ex.1002**), the PubMed records for *Leary* (**Appendix 1002A**), and the *Science Translational Medicine* website archived by the Internet Archive (**Appendix 1002B**) inform my opinion that *Leary* was published online on February 24, 2010, highlighted as a research article on the journal website on February 25, 2010, indexed by PubMed and became discoverable in PubMed by June 16, 2010. Based on my knowledge and experience with online databases, it is my opinion that *Leary* would have been publicly accessible by February 25, 2010, on the Internet, and no later than June 16, 2010,

when it became discoverable in PubMed. Citation history shows the earliest citation appeared in March 2010.

III. Authenticity and Public Availability of *Chan* (Ex.1008)

A. Authentication

23. **Ex.1008** is a true and correct copy of Cancer genome scanning in plasma: detection of tumor-associated copy number aberrations, single-nucleotide variants, and tumoral heterogeneity by massively parallel sequencing, (“*Chan*”), by Chan et al., *Clinical Chemistry*, vol. 59, no. 1, pp. 211-224, that I obtained from the publisher. I began my research on Google Scholar by searching by the title of *Chan* for records, and a retrieved record led me to the publisher’s website for *Chan* at <https://academic.oup.com/clinchem/article/59/1/211/5622098>. I then followed the PDF link to acquire a copy. This copy from Elsevier is presented as **Ex.1008** in this declaration.

24. **Ex.1008** is a true and correct copy of *Chan* obtained from the publisher. Page 1 (internal page 211) shows the title, authors, and a “2012” copyright date for *Chan*, with the American Association for Clinical Chemistry as the copyright holder. The top of the page shows *Chan* appears in *Clinical Chemistry* “59:1, 211-244 (2013).” The bottom of the page shows the manuscript was received on September 6, 2012 and accepted on September 27, 2012, and this article was “[p]reviously published online at DOI: 10.1373/clinchem.2012.196014.” **Ex.1008** shows *Chan*

has a total of 14 pages, including four color figures, three tables, and 35 references.

25. The publisher's copy of *Chan* (**Ex.1008**) is in a condition that creates no suspicion about its authenticity. Specifically, this copy is not missing any intermediate pages, the text on each page appears to flow seamlessly from one page to the next, and there are no visible alterations to the document. Moreover, **Ex.1008** is in the custody of the publisher, a place where, if authentic, it would likely be found. Accordingly, I see no reason to question the authenticity of this copy of *Chan* (**Ex.1008**).

B. PubMed Record for *Chan*

26. To determine whether *Chan* was indexed, I searched PubMed by the article title for records and located a record at <https://pubmed.ncbi.nlm.nih.gov/23065472/>. This PubMed record is presented as **Appendix 1008A** in this declaration. PubMed is an authoritative system that contains “more than 34 million citations for biomedical literature from MEDLINE, life science journals, and online books” and PubMed records may include links to “full text content from PubMed Central and publisher web sites” (source: <https://pubmed.ncbi.nlm.nih.gov>). I personally located, identified and obtained the PubMed metadata record for *Chan*. I saved the metadata record in PDF and also saved the record in PubMed format to show the record details in data fields.

27. Appendix 1008A is a true and correct copy of the PubMed metadata

record for *Chan*. The top of page 1 shows a full text link to “Oxford Academic” that takes users to the publisher’s website for *Chan*. Page 1 also shows the title, authors, and an abstract that covers the background, methods, results, and conclusions of this study. A note after the abstract shows *Chan* has a “2012” copyright date, with the American Association for Clinical Chemistry as the copyright holder. The area above the title shows a note of “Clin Chem. 2013 Jan; 59(1):211-24. Doi: 10.1373/clinchem.2012.196014. Epub 2012 Oct 11” that indicates *Chan* was first published online in *Clinical Chemistry* on October 11, 2012, with a doi (digital object identifier) of “10.1373/clinchem.2012.196014.” This note is consistent with the note on page 1 of **Ex.1008** that *Chan* was previously “published online at DOI: 10.1373/clinchem.2012.196014.” This note also indicates that *Chan* was later published in the print version of *Clinical Chemistry*, vol. 59, no. 1, pages 211 to 244, and this print issue has a “January 2013” issue date. The PMID (PubMed Identifier) for this metadata record is “23065472.” The authors, title, copyright date, copyright holder, journal title, issue numbering, page numbers and doi match the information presented in the publisher’s copy of *Chan* (**Ex.1008**), confirming this record represents *Chan*.

28. The “Comment in” note area following the abstract shows C. Swanton commented on *Chan* in an article entitled “Plasma-derived tumor DNA analysis at whole-genome resolution”, which was published online on October 31, 2012, and

later appeared in the print issue of *Clinical Chemistry*, vol. 51, issue 1, pages 6 to 8, and the print issue has a “January 2013” issue date. The PubMed metadata record for *Chan* (**Appendix 1008A**) informs my opinion that *Chan* was published online on October 11, 2012, and Swanton’s commentary on *Chan* was published online on October 31, 2012. Both articles were later published in the print version of *Clinical Chemistry* vol. 59, no. 1.

29. Pages 3 to 4 of **Appendix 1008A** are the PubMed record in PubMed format to show data in specific fields. I use the “MEDLINE/PubMed Data Element (Field) Descriptions” (<https://www.nlm.nih.gov/bsd/mms/medlineelements.html>) to interpret the record. The “PMID” field shows “23065472” as the PubMed identifier for the record for *Chan*. The “STAT” field shows the status of the record is “MEDLINE” (meaning Medical subject headings were assigned to the article), and the DCOM field shows this index record was completed on “20130429” (*i.e.*, April 29, 2013). The first IS field shows the ISSN (International Standard Serial Number) for the electronic version of this journal is “1530-8561” and the ISSN for the print journal is “0009-9147.” Fields VI (volume), IP (issue), DP (date of publication), Ti (title), PG (pages) and JT (journal title) show *Chan* was published in vol. 59, no. 1 (January 2013) of *Clinical Chemistry* from pages 211 to 224. The first FAU (full author) field shows “Chan, K C Allen” as the first author. Field “DEP” (date of electronic publication) on page 4 of **Appendix 1008A** shows *Chan* was published

electronically on “20121011” (*i.e.*, October 11, 2012). Field CRDT (creation date) and Field EDAT (entrez date) show the metadata record was added to PubMed on “2012/10/16” (*i.e.*, October 16, 2012), and Field MHDT (MeSH date) shows medical subject headings were added to the record on “2013/04/30” (*i.e.*, April 30, 2013) and the record was elevated to MEDLINE (source: <https://pubmed.ncbi.nlm.nih.gov/help/>). The PubMed record shows nine medical subject headings were assigned to indicate the subjects of the *Chan* article. Field “AID” (Article identifier) shows the doi of *Chan*. Field “CIN” (Comment in) shows an article containing a commentary on *Chan* appears in “Clin Chem. 2013 Jan; 59 (1): 6-8. PMID: 23115052.” The PMID (PubMed Identifier) refers to the Swanton commentary (see page 1 of **Appendix 1008A**) that was first published electronically on October 31, 2012.

30. This PubMed metadata record (**Appendix 1008A**) made *Chan* discoverable in PubMed on October 16, 2012, after *Chan* was published electronically on October 11, 2012. Interested users would have been able to search for *Chan* in PubMed by the authors, title, journal title, ISSN, and abstract. Medical subject headings were added to enrich the PubMed record on April 30, 2013. From that date on, interested users would have been able to conduct subject searches to discover *Chan*.

C. Swanton Commentary on *Chan*

31. The PubMed metadata record for *Chan* shows Swanton commented on *Chan* and the Swanton commentary was published electronically on October 31, 2012, and indexed by PubMed metadata record 23115052. I retrieved this PubMed metadata record and followed the full text link to obtain a copy of *Swanton* from the publisher. This publisher copy of *Swanton* is presented as **Appendix 1008B**.

32. **Appendix 1008B** is a true and correct copy of Plasma-derived tumor DNA analysis at whole-genome resolution, (“*Swanton*”), by Charles Swanton that I obtained from the publisher. Page 1 shows the title and author and indicates this article is an editorial. The top of this page shows *Swanton* appears in “Clinical Chemistry 59:1, 6-8 (2013)” meaning vol. 59, no. 1 (2013), pages 6 to 8. The publication history at the bottom of page 1 shows the manuscript was received on “October 2, 2012; accepted October 4, 2012” and *Swanton* was “[p]reviously published online at DOI: 10.1373/clinchem.2012.197053.” **Appendix 1008B** shows *Swanton* has a total of three pages, including 16 references, and reference 16 is the *Chan* article.

33. *Swanton* was also indexed by PubMed (PMID 23115052). **Appendix 1008C** is a true and correct copy of the PubMed metadata record for *Swanton*. I personally located, identified and obtained this record from PubMed. This is the type of record experts in my field would reasonably rely upon when forming their

opinions.

34. The top of the PubMed Metadata record (**Appendix 1008C**) shows a full text link to “Oxford Academic” that takes users to the publisher’s webpage for *Swanton* at <https://academic.oup.com/clinchem/article/59/1/6/5622134?login=false>. The metadata record shows the title and author of *Swanton*, the PMID (PubMed identifier) for this record as “23115052” and the DOI (digital object identifier) of *Swanton* as “10.1373/clinchem.2012.197053.” The area above the title shows a note of “Editorial > Clin Chem. 2013 Jan; 59(1): 6-8. doi: 10.1373/clinchem.2012.197053. Epub 2012 Oct 31” that indicates *Swanton* is an editorial that was first published online in *Clinical Chemistry* on October 31, 2012, and later appeared in the print version of *Clinical Chemistry*, vol. 59, no. 1, pages 6 to 8, and the print issue has an issue date of January 2013. The “Comment on” area shows *Swanton* contains a commentary on *Chan*, which was published electronically on October 11, 2012.

35. Page 2 of **Appendix 1008C** is the PubMed record saved in PubMed format to show data in specific fields. I use the “MEDLINE/PubMed Data Element (Field) Descriptions” (<https://www.nlm.nih.gov/bsd/mms/medlineelements.html>) to interpret the record. The “PMID” field shows “23115052” as the PubMed identifier, and the DCOM field shows this index record was completed on “20130429” (*i.e.*, April 29, 2013). The first IS field shows the ISSN (International Standard Serial

Number) for the electronic version of this journal is “1530-8561” and the ISSN for the print journal is “0009-9147.” Fields VI (volume), IP (issue), DP (date of publication), Ti (title), PG (pages) and JT (journal title) show *Swanton* was published in vol. 59 no. 1 (January 2013) of *Clinical Chemistry* from pages 6 to 8. The FAU (full author) field shows “Swanton, Charles” as the author. Field “DEP” (Date of electronic publication) shows *Swanton* was published electronically on “20121031” (*i.e.*, October 31, 2012). Field CRDT (Creation date) and Field EDAT (Entrez date) show the metadata record was added to PubMed on “2012/11/02” (*i.e.*, November 2, 2012) and Field MHDT (MeSH date) shows medical subject headings were added to the record on “2013/04/30” (*i.e.*, April 30, 2013) to elevate this record into MEDLINE (source: <https://pubmed.ncbi.nlm.nih.gov/help/>). Six medical subject headings were assigned to the record to indicate the subjects of *Swanton*. Field “AID” (Article identifier) shows the doi of *Swanton*. Field “CON” (Comment on) shows the work commented on by Swanton appears in “Clin Chem. 2013 Jan; 59(1) : 211-24. PMID: 23065472.” This PMID refers to *Chan* (see page 1), which was first published electronically on October 11, 2012.

36. This PubMed metadata record (**Appendix 1008C**) clarifies the relationship between *Swanton* and *Chan* and shows that *Swanton* became discoverable in PubMed on November 2, 2012, after it was published on October 31, 2012. As of November 2, 2012, interested users would have been able to search

for *Swanton* in PubMed by title, author, journal title, and ISSNs. They would also have been able to use the note in the “Common on” note area to discover *Chan*. Six medical subject headings were added to enrich the PubMed record for *Swanton* on April 29, 2013, enabling users to conduct subject searches to locate *Swanton*.

D. Copyright registration record

37. Appendix 1008D is a true and correct copy of the copyright registration record for vol. 59, no.1 of *Clinical Chemistry* that contains *Chan* and *Swanton*. I obtained this record from the public catalog of the United States Copyright Office in the Library of Congress by first accessing the <http://www.copyright.gov/> website, selecting the “Learn More” link under “Search Copyright Records” and following the link for “Official Public Catalog (1978 to present)” under “Online Records Collections” to perform a journal title search to retrieve the record. I personally located, identified and obtained this record. This is the type of record experts in my field would reasonably rely upon when forming their opinions.

38. The “Type of Work” field identifies this publication as “Serial”. The title and copyright claimant match the information presented in **Ex.1008**. The “Issues Registered” field shows “vol. 59, no. 1, January 2013. Created 2012; Pub. 2012-12-31. Reg. 2013-01-31; TX0007648861” meaning vol. 59, no. 1 was published on December 31, 2012, and has a copyright registration number of TX0007648861.

E. Summary of My Opinion on *Chan*

39. Taken together, the publisher's copy of *Chan* (**Ex.1008**), the PubMed records for *Chan* (**Appendix 1008A**), the publisher's copy of *Swanton* (**Appendix 100B**), and the PubMed record for *Swanton* (**Appendix 1008C**) inform my opinion that *Chan* was published online on October 11, 2012, and became discoverable on PubMed on October 16, 2012, and that the Swanton commentary on *Chan* was published online on October 31, 2012, and became discoverable on PubMed on November 2, 2012. Both online publications were later published in the print version of vol. 59, no. 1 of *Clinical Chemistry* that has a "January 2013" issue date, and the copyright registration record shows this issue was published on December 31, 2012. It is therefore my conclusion that *Chan* was published on October 11, 2012, and indexed by PubMed on October 16, 2012. From that date on, interested users would have been able to discover *Chan* on PubMed to obtain the full text of this article.

IV. Conclusion

40. I declare under penalty of perjury that the foregoing is true and correct.

Executed on Date: 11-17-2022


Ingrid Hsieh-Yee, Ph.D.

Ingrid Hsieh-Yee
Professor (retired)
Dept. of Library and Information Science
Catholic University of America
Washington, D.C. 20064
E-mail: hsiehyee@cua.edu

Education

Ph.D. 1990. Library and Information Studies, University of Wisconsin-Madison
Minors: Sociology and Psychology

M.A. 1984. Library and Information Studies, University of Wisconsin-Madison.

M.A. 1983. Comparative Literature, University of Wisconsin-Madison.

B.A. 1980. Foreign Languages and Literature, National Taiwan University.

Work Experience

Professor, School/Dept. of Library and Information Science, Catholic University of America,
2004-2022, retired in August 2022 (Assistant Professor, 1990-1996; Associate Professor,
1997-2004)

Co-Chair, Dept. of Library and Information Science, Catholic University of America, June 2015-
August 2016.

Acting Dean, School of Library and Information Science, Catholic University of America,
January 2010-June 2012.

Cataloger, Dept. of Legislative Reference Library, Annapolis, Maryland, 1989-1990.

Lecturer, School of Library and Information Studies, University of Wisconsin-Madison, 1988.

Teaching Assistant, School of Library and Information Studies, University of Wisconsin-
Madison, 1986-1988.

Cataloger, Health Sciences Library, University of Wisconsin-Madison, 1984-1986.

Areas of Teaching and Research Interests

Information Organization and Access; Metadata; Cataloging & Classification; Information
Architecture; Information Retrieval; Digital Collections; Social Media; Scholarly

Communication; Information Behavior; Health Informatics; Human Computer Interaction; Usability Studies

Grants & Honors

Cultural Heritage Information Management Project. IMLS grant. Amount: \$498,741. Period: Aug. 2012 to July 2015. Co-PI with Dr. Youngok Choi.

D.C. Health Information Technology (HIT4): Building Capacity & Providing Access in Our Nation's Capital. Dept. of Labor H2B Training Grant. Grant amount: \$4,175,500. Grant period: Nov. 2011 to Dec. 2015. Partner with the Metropolitan School of Professional Studies of the Catholic University of America, Children's National Medical Center, D.C. Department of Employment Services, Holy Cross Hospital, Howard University, Center for Urban Progress, Providence Hospital, and Sibley Memorial Hospital.

Capital Health Careers Project. Department of Labor Healthcare Sector and Other High Growth and Emerging Industries Grant. Grant amount: \$4,953,999. Grant period: March 2010 – February 2013. Awarded to a group of healthcare organizations and educational institutions in Washington, D.C. Providence Health Foundation of Providence Hospital (Lead institution). Part of the grant supported the development of a Master's degree program in Information Technology with a concentration in Health Information Technology offered by the School of Library and Information Science.

The Washington D.C. School Librarians Project. IMLS grant. Grant amount: \$412,660. Grant period: Aug. 2007 – June 2011. The School partnered with the District of Columbia Public Schools (DCPS) and the District of Columbia Library Association to educate and mentor school media specialists for the DCPS system. PI, Jan. 2010 to June 2011.

SIG Member of the Year, American Society for Information Science and Technology (2009).

Most Outstanding Paper of *OCLC Systems & Services* (2001).

ALISE Research Grant (2001).

Most Outstanding Paper of *OCLC Systems & Services* (2000).

Research Grant from ERIC (1999-2000).

Best Research Paper Award; Association for Library and Information Science Education (1998).

Research Grants, Catholic University of America. 1991, 1992, 1993, 1996, 1998, 1999, 2004, 2005, 2006, 2007, 2013-14.

Cooperative Faculty Research Grant, Consortium of Universities in the Washington Metropolitan Area (1993-1994).

Cooperative Research Grant, Council on Library Resources (1993-1994).

Journal of the American Society for Information Science Best Paper Award (1993).

ASIS/ISI Information Science Doctoral Dissertation Scholarship (1989).

HEA Title IIB Fellowship (Dept. of Education) (1989)

Chinese-American Librarians Association Scholarship (1987).

Beta Phi Mu (1985).

Vilas Fellowship, University of Wisconsin-Madison. 1984

Publications

Hsieh-Yee, I. (2021). Can We Trust Social Media? *Internet Reference Services Quarterly*, 1-15.
Published online: 28 July, 2021. 10.1080/10875301.2021.1947433.

Bailey, T., & Hsieh-Yee, I. (2020). Combating the Sharing of False Information: History, Framework & Literacy Strategies, *Internet Reference Services Quarterly*, 24 (1-2), 9-30.

Choi, Y., and Hsieh-Yee, I. (2010). Finding Images in an OPAC: Analysis of User Queries, Subject Headings, and Description Notes. *Canadian Journal of Information and Library Science*, 34(3): 271 – 295.

Hsieh-Yee, I. (2008). Educating Cataloging Professionals in a Changing Information Environment. *Journal of Education for Library and Information Science*, 46(2): 93-106.

Vellucci, S. L., Hsieh-Yee, I., and Moen, W.E. (2007). The Metadata Education and Research Information Commons (MERIC): A Collaborative Teaching and Research Initiative. *Education for Information*, 25(3&4): 169-178.

NISO Framework for Guidelines for Building Good Digital Collections. 3rd ed. Baltimore, MD: National Information Standards Organization, 2007. Also available online: <http://www.niso.org/framework/framework3.pdf> (NISO Working Group members: Priscilla Caplan (chair), Grace Agnew, Murtha Baca, Tony Gill, Carl Fleischhauer, Ingrid Hsieh-Yee, Jill Koelling, and Christie Stephenson.)

Choi, Y., Hsieh-Yee, I., and Kules, B. (2007). Retrieval Effectiveness of TOC and LCSH. *Proceedings of the Joint Conference on Digital Libraries*, pp. 233-234.

Vellucci, S., and Hsieh-Yee, I. (2007). They Didn't Teach Me That in Library School! Building

a Digital Teaching Commons to Enhance Metadata Teaching, Learning and Research. *Proceedings of the National Conference of the Association of College and Research Libraries, Baltimore, MD*, pp. 26-31.

Mitchell, Vanessa, and Ingrid Hsieh-Yee. (2007). Converting Ulrich's Subject Headings to FAST Headings: A Feasibility Study. *Cataloging & Classification Quarterly*, 45(1): 59-85.

Hsieh-Yee, I., Tang, R., and Zhang, S. (2007). User Perceptions of a Federated Search System. *IEEE Technical Committee on Digital Libraries Bulletin*, Summer 3(2) (URL = <http://www.ieee-tcdl.org>).

Tang, R., Hsieh-Yee, I., and Zhang, S. (2007). User Perceptions of MetaLib Combined Search: An Investigation of How Users Make Sense of Federated Searching." *Internet Reference Services Quarterly*, 12(12): 211-236.

Hsieh-Yee, I., Tang, R., and Zhang, S. (2006). User Perceptions of a Federated Search System. *Proceedings of the Joint Conference on Digital Libraries, June 11-15, 2006, Chapel Hill*, p. 338.

Hsieh-Yee, I. (2006). *Organizing Audiovisual and Electronic Resources for Access: A Cataloging Guide*. 2nd ed. Westport, Conn.: Libraries Unlimited.

NISO A Framework of Guidance for Building Good Digital Collections. 2nd ed. Bethesda, MD: National Information Standards Organization, 2004. Framework Advisory Group: Grace Agnew, Liz Bishoff, Priscilla Caplan (Chair), Rebecca Gunther and Ingrid Hsieh-Yee.

Hsieh-Yee, I. (2004). Cataloging and Metadata Education in North American LIS Programs. *Library Resources & Technical Services*, 48(1): 59-68.

Hsieh-Yee, I. (2004). Cataloging and Metadata Education. In Gary E. Gorman (Ed.), *International Yearbook of Library and Information Management 2003: Metadata Applications and Management*, (pp.204-234). London: Facet Publishing.

Yee, P. L., Hsieh-Yee, I., Pierce, G.R., Grome, R., and Schantz, L. (2004). Self-Evaluative Intrusive Thoughts Impede Successful Searching on the Internet. *Computers in Human Behavior*, 20(1): 85-101.

Hsieh-Yee, I. (2003). Cataloging and Metadata Education: A Proposal for Preparing Cataloging Professionals of the 21st Century. A report submitted to the ALCTS-Education Task Force in response to Action Item 5.1 of the *Bibliographic control of Web Resources: A Library of Congress Action Plan*. Approved by the Association for Library Collections and Technical Services. Web version available since April 2003 at <http://lcweb.loc.gov/catdir/bibcontrol/CatalogingandMetadataEducation.pdf>.

- Hsieh-Yee, I. (2002). Cataloging and Metadata Education: Asserting a Central Role in Information Organization. *Cataloging & Classification Quarterly* 34(½): 203-222.
- Hsieh-Yee, I., and Smith, M. (2001). The CORC Experience: Survey of Founding Libraries, Part I. *OCLC Systems & Services*, 17: 133-140. (Received "The Most Outstanding Paper of OCLC Systems & Services in 2001" award.)
- Hsieh-Yee, I., and Smith, M. (2001). The CORC Experience: Survey of Founding Libraries, Part II, Automated Tools and Usage. *OCLC Systems & Services*, 17: 166-177. (Received "The Most Outstanding Paper of OCLC Systems & Services in 2001" award.)
- Hsieh-Yee, I. (2001). ERIC User Services: Changes and Evaluation for the Future. *Government Information Quarterly*, 18: 31-42.
- Hsieh-Yee, Ingrid. (2001). Research on Web Search Behavior. *Library and Information Science Research*, 23: 167-185.
- Logan, E., and Hsieh-Yee, I. (2001). Library and Information Science Education in the Nineties. *Annual Review of Information Science and Technology*, 35: 425-477.
- Hsieh-Yee, I. (Ed.) (2001). *Library and Information Science Research*, 23 (2). A special issue in honor of the retirement of Douglas L. Zweizig.
- Hsieh-Yee, I. (2000). *ERIC User Services: Evaluation in a Decentralized Environment*. Washington, D.C.: Dept. of Education.
- Hsieh-Yee, Ingrid. (2000). *Organizing Audiovisual and Electronic Resources for Access: A Cataloging Guide*. Littleton, CO: Libraries Unlimited.
- Hsieh-Yee, I. (2000). Organizing Internet Resources: Teaching Cataloging Standards and Beyond. *OCLC Systems & Services*, 16: 130-143. (Received "The Most Outstanding Paper of OCLC Systems & Services in 2000" award.)
- Hsieh-Yee, I. (1998). The Retrieval Power of Selected Search Engines: How Well Do They Address General Reference Questions and Subject Questions? *Reference Librarian*, 60: 27-47.
- Hsieh-Yee, I. (1998). Search Tactics of Web Users in Searching for Texts, Graphics, Known Items and Subjects: A Search Simulation Study. *Reference Librarian*, 60: 61-85. (Received the 1997 Best ALISE Research Paper Award.)
- Hsieh-Yee, I. (1997). Access to OCLC and Internet Resources: LIS Educators' Views and Teaching Practices. *RQ*, 36: 569-86.

- Hsieh-Yee, I. (1997). Teaching Online and CD-ROM Resources: LIS Educators' Views and Practices. *Journal of Education for Library and Information Science*, 38: 14-34.
- Hsieh-Yee, I. (1996). The Cataloging Practices of Special Libraries and Their Relationship with OCLC. *Special Libraries*, 87: 10-20.
- Hsieh-Yee, I. (1996). Modifying Cataloging Practice and OCLC Infrastructure for Effective Organization of Internet Resources. In *Proceedings of the OCLC Internet Cataloging Colloquium*. [Online]. Available: <http://www.oclc.org/oclc/man/colloq/hsieh.htm>
- Hsieh-Yee, I. (1996). Student Use of Online Catalogs and Other Information Channels. *College & Research Libraries*, 57: 161-175.
- Hsieh-Yee, I. (1995). Ten entries in James S. C. Hu (Ed.), *Encyclopedia of Library & Information Science*, 913, 1028-29, 1036, 1037, 1145-46, 1514, 1575, 1763-64, 2216-27, 2378-79. Taipei, Taiwan: Sino-American Publishing. (Topics include "Advanced Technology/Libraries," "Information Ethics," "Instruction on Cataloging and Classification," "Instruction on Reference Services.")
- Hsieh-Yee, I. (1993). Effects of Search Experience and Subject Knowledge on Online Search Behavior: Measuring the Search Tactics of Novice and Experienced Searchers. *Journal of the American Society for Information Science*, 44: 161-174. (Received the 1993 Best JASIS Paper Award.)

Works in Progress

- Hsieh-Yee, I. Actionable lessons for debunking false information: A comparative analysis of literacy lessons in communication and library and information science
- Hsieh-Yee, I. Representing diverse voices and communities in cataloging and indexing: The evolution of controlled vocabularies
- Hsieh-Yee, I. Information Organization and Retrieval in Practice (a monograph scheduled for completion and publication in 2023)

Presentations

- Bailey, T. & Hsieh-Yee, I. (February 2020). The Phenomena of Sharing Misinformation and the Need for Information Literacy. Presented at the 2020 Bridging the Spectrum Symposium, Washington, D.C., February 14, 2020.

- Hsieh-Yee, I. and Fragan-Fly, J. (May 2018) Trends, Design & Strategies for Digital Scholarship Services. Presented at the 2018 Maryland/Delaware Library Association Conference, Cambridge, MD.
- Hsieh-Yee, I. (February, 2018) Research Data Management: What It Takes to Succeed. Presented at the 10th Bridging the Spectrum Symposium, Washington, D.C.
- Hsieh-Yee, I. (February, 2017) *Research Data Management: New Competencies and Opportunities for Information Professionals*. Presented at the 9th Bridging the Spectrum Symposium, Washington, D.C.
- Hsieh-Yee, I. and Lawton, P. (February, 2017) *Enhancing Catholic Portal Searches with User Terms and LCSH*. Presented at the 9th Bridging the Spectrum Symposium, Washington, D.C.
- Hsieh-Yee, I. (2016, October) *Visualizing Data for Information*. Presented at the 2016 Virginia Library Association Conference, Hot Springs, VA.
- Hsieh-Yee, I. (2016, August) *Religious Materials Toolbox for Archivists: Solutions to Problems Facing the Profession*. Presented at Archives * Records 2016, Atlanta, GA.
- Hsieh-Yee, I. and Lawton, P. (2016, March) *Enhancing Retrieval of Catholic Materials with LCSH Knowledge Structure*. Presented at the 2016 Catholic Library Association Conference, San Diego, CA.
- Fagan-Fry, J. and Hsieh-Yee, I. (2016, February) *Approaches to Digital Scholarship at Top Universities around the World: Scholarly Publishing in the Digital Age*. Presented at the 8th Bridging the Spectrum Symposium, Washington, D.C.
- Hsieh-Yee and Fagan-Fry, J. (2016, January) *Innovative Services for Digital Scholarship at Top 100 Research Libraries of the World*. Poster presented at the 2016 Annual Conference of the Association for Library and Information Science Education, Boston, Mass.
- Hsieh-Yee, I. and Lawton, P. (2015, June). *Crowdsourcing terms for CRRA portal themes*. Poster presented at the third CRRA symposium and annual meeting, Bringing the created toward the Creator: Liturgical art and design since Vatican II. Catholic Theological Union, Chicago, Illinois.
- Hsieh-Yee, I. and Lawton, P. (2015, February). *Crowdsourcing terms for thematic exploration in the Catholic Portal*. Poster presented at the 7th Annual Bridging the Spectrum Symposium, Washington, D.C.
- Hsieh-Yee, I., James, R., and Fagan-Fry, J. (2015, February). *Support for digital scholarship at top university libraries of the world*. Poster presented at the 7th Annual Bridging the Spectrum Symposium, Washington, D.C.

- Hsieh-Yee, I., Zhang, S., Lin, K., and Cherry, S. (2015, February). *Thus said the end users: Summon experience and support for research workflows*. Poster presented at the 7th Annual Bridging the Spectrum Symposium, Washington, D.C.
- Yontz, E., Hsieh-Yee, I., & Houston, S. (2015, February). *Healthy Heroes Summer Reading Club: Developing healthy youth at public libraries*. 11th Annual Jean Mills Health Symposium, Greenville, North Carolina.
- Yontz, E., Hsieh-Yee, I., and Houston, S. (2015, January). *Healthy youth and libraries: A pilot study*. Association for Library & Information Science Education (ALISE) Annual Conference, Chicago, Illinois.
- Hsieh-Yee, I. (2014, May). *Linking CRRRA resources to portal themes via authority files*. Presented at the Catholic Research Resources Alliance 2014 Membership Meeting, Marquette, WI.
- Hsieh-Yee, I. (2014, April). *Enhancing subject access to CRRRA resources*. Presented at the 2014 Catholic Library Association Conference, Pittsburgh, PA.
- Hsieh-Yee, I. (2014, January). *Health Information Technology Program: Educational entrepreneurship in action*. Presented at the 2014 annual Conference of the Association for Library and Information Science Education, Philadelphia, PA.
- Hsieh-Yee, I., Zhang, S., Lin, K., and Cherry, S. (2014, January). *Discovering information through Summon: An analysis of user search strategies and search success*. Paper presented at the 6th Bridging the Spectrum Symposium, Washington, D.C.
- Hsieh-Yee, I. (2012, December). *National Digital Stewardship Alliance and SLIS at CUA: An Educational Partnership*. Paper presented at Best Practices Exchange: Acquiring, Preserving, and Providing Access to Government Information in the Digital Era, Annapolis, MD.
- Choi, Y. and Hsieh-Yee, I. (2010, January). *Finding Images in an OPAC: Analysis of User Queries, Subject Headings, and Description Notes*. Paper presented at 2nd Annual Bridging the Spectrum Symposium, Catholic University of America, Washington, D.C.
- Hsieh-Yee, I. and Coogan, J. (2010, January). *Google Scholar vs. Academic Search Premier: What Libraries and Searchers Need to Know*. Paper presented at 2nd Annual Bridging the Spectrum Symposium, Catholic University of America, Washington, D.C.
- Hsieh-Yee, I. (2009, November). *Information Science Education: An LIS School's Perspective*. Paper presented at Annual Meeting of the American Society for Information Science and Technology, Vancouver, British Columbia, Canada.

- Hsieh-Yee, I., Menard, E., Ya-Ning Chen, A., Shu-Jiun Chen, S., Kalfatovic, M. R., Wisser, K. M. (2009, November). *Information Organization in Libraries, Archives and Museums: Converging Practices and Collaboration Opportunities*. Presented at Annual Meeting of the American Society for Information Science and Technology, Vancouver, British Columbia, Canada. (Organizer and moderator of this panel.)
- Hsieh-Yee, I. and Coogan, J. (2009, July). *Catching up to Google Scholar: The Retrieval Power of Academic Search Premier and Google Scholar*. Poster presented at American Library Association Conference, Chicago, Illinois.
- Hsieh-Yee, I., with the CUA Scholarly Communications Project Team. (2009, January). *Digital Scholarship@CUA: Developing an Institutional Repository for CUA*. Poster presented at 1st Annual Bridging the Spectrum Symposium, Catholic University of America, Washington, D.C.
- Wise, M., Cylke, K., and Hsieh-Yee, I. (2009, January). *Digital Talking Books: Meeting the Needs of the Blind and the Handicapped*. Paper presented at the Bridging the Spectrum Symposium, Catholic University of America, Washington, D.C.
- Hsieh-Yee, I. (2009, January). *User Expectations of MERIC*. Presented at the Information Organization Competencies for the 21st Century Discussion Session of the 2009 Conference of the Association for Library and Information Science Education, Denver, Colorado.
- Choi, Y., and Hsieh-Yee, I. (2008, November). *Subject Access for Images in an OPAC*. Annual Meeting of the American Society for Information Science and Technology, Columbus, Ohio. (Also co-organized a panel on Retrieving and Using Visual Resources: Challenges and Opportunities for Research and Education.)
- Hsieh-Yee, I. (2008, June). *Educating Cataloging Professionals in a Changing Information Environment*. National Taiwan University, Taipei, Taiwan.
- Vellucci, S. L., Moen, W.E., Hsieh-Yee, I., Marson, B., and Wisser, K. (2008, January) *Building a Metadata Education and Research Community through MERIC (Metadata Education and Research Information Commons): Demo and Stakeholder Input*. A panel presented at the 2008 Conference of the Association for Library and Information Science Education, Philadelphia, Pennsylvania.
- Hsieh-Yee, I., Choi, Y. and Kules, B. (2007, October). *Searching for Books and Images in OPAC: Effects of LCSH, TOC and Subject Domains*. A poster presented at the American Society for Information Science and Technology Annual Meeting, Milwaukee, Wisconsin.
- Hsieh-Yee, I. and Coogan, J. (2007, August) *Google Scholar vs. Academic Search Premier: A Comparative Analysis*. Presented to the Faculty and Staff of the University of the District

of Columbia.

- Hsieh-Yee, I. and Coogan, J. (2007, June). *Google Scholar vs. Academic Search Premier: A Comparative Analysis*. Presented to the Washington Research Library Consortium Community, Catholic University of America, Washington, D.C.
- Hsieh-Yee, I., Choi, Y., and Kules, B.. (2007, June). *What Users Need for Subject Access: Table of Contents or Subject Headings?* A poster presented at the 2007 American Library Association Annual Conference, Washington, D.C., June 2007.
- Choi, Y., Hsieh-Yee, I., and Kules, B. (2007, June). *Retrieval Effectiveness of TOC and LCSH*. A paper presented at the Joint Conference on Digital Libraries 2007, Vancouver, Canada.
- Vellucci, S. L., Hsieh-Yee, I., and Moen, W.E. (2007, May). *If We Build It, Will They Come? Building a Community of Practice for Metadata Stakeholders*. A poster presented at the Rutgers University Research Day, Bridgeton, New Jersey.
- Hsieh-Yee, I. (2007, May). *Federated Searching: User Experience & Perceptions*. International Conference on Information Organization & Retrieval, National Taiwan University, Taipei, Taiwan.
- Hsieh-Yee, I. (2007, May). *Search Performance of Google Scholar and Academic Search Premier*. International Conference on Information Organization & Retrieval, National Taiwan University, Taipei, Taiwan.
- Hsieh-Yee, I. (2007, May) *MERIC: Building a Digital Commons for Metadata Education & Research*. International Conference on Information Organization & Retrieval, National Taiwan University, Taipei, Taiwan.
- Hsieh-Yee, I., and Coogan, J. (2007, March/April). *A Comparative Analysis of Google Scholar and Academic Search Premier*. Poster presented at the Association of College & Research Libraries 13th National Conference, Baltimore, Maryland.
- Vellucci, S. L. and Hsieh-Yee, I. (2007, March/April) *They Didn't Teach Me That in Library School! Building a Digital Teaching Commons to Enhance Metadata Teaching, Learning and Research*. On-site presentation and Webcast by Elluminate. A contributed paper presented at the Association of College & Research Libraries 13th National Conference, Baltimore, Maryland. The acceptance rate for contributed paper was 20%. This paper was one of 10 conference papers chosen for live webcast during the conference.
- Moen, W., Hsieh-Yee, I. and Vellucci, S.L. (2007, January) *A DSpace Foundation for a Teaching & Research Commons: The Metadata Education and Research Information Commons*. A poster session presented at the Open Repositories Conference 2007, San Antonio, Texas.

- Tang, R., Hsieh-Yee, I., and Zhang, S. (2006, November) *User Perception of MetaLib Combined Search*. Paper presented at the Annual Meeting of the American Society for Information Science and Technology, Austin, Texas, Nov. 2006.
- Hsieh-Yee, I. (2006, November). *Federated Searching: User Perceptions, System Design, and Library Instructions*. Paper presented at the Annual Meeting of the American Society for Information Science and Technology, Austin, Texas. (Panel organizer, moderator, presenter).
- Hsieh-Yee, I. (2006, November). *Building a Digital Teaching Commons to Enhance Teaching and Learning: The MERIC Experience and Challenges*. Paper presented at the Annual Meeting of the American Society for Information Science and Technology, Austin, Texas. (Panel organizer, moderator, presenter)
- Hsieh-Yee, I. (2006, September). *Search Performance of Google Scholar and Academic Search Premier*. Paper presented at the ERIC Publishers Meeting, Washington, D.C.
- Hsieh-Yee, I., Zhang, S., and Rong Tang, R. (2006, June). *User Perceptions of a Federated Search System*. Poster presented at Joint Conference on Digital Libraries, Chapel Hill, North Carolina.
- Hsieh-Yee, I. and Zhang, S. (2006, June). *Preparing Users for Federated Search: Implications of a MetaLib User Perceptions Study*. Paper presented at the 2006 Ex Libris User Groups of North America Conference, Knoxville, Tennessee.
- Hsieh-Yee, I. (2006, January). *MERIC Organizations and Navigation*. Paper presented at the 2006 ALISE Annual Conference, San Antonio, Texas.
- Hsieh-Yee, I. (2006, January). *Metadata and Cataloging Education: Recommended Competencies*. Paper presented at the 2006 ALISE Annual Conference, San Antonio, Texas.
- Hsieh-Yee, I. (2005, November). *Digital Library Evaluation: Progress & Next Steps*. Presentation at the Annual Meeting of the American Society for Information Science & Technology, Charlotte, North Carolina.
- Hsieh-Yee, I. (2005, August). *Providing Access to Digital Content: Issues for DL Managers*. Presentation at MDK12 Digital Library Steering Committee Meeting, Columbia, Maryland.
- Hsieh-Yee, I. (2005, April). *Enhancing Teaching and Learning: The Role of School Library Media Specialists*. Presentation at Meeting of the Baltimore County Public School System School Media Specialists, Baltimore, Maryland.

- Hsieh-Yee, I. (2005, January). *Subject Access and Users: Insights & Inspirations from Marcia J. Bates*. Paper presented at the Historical Perspectives SIG, 2005 Conference of the Association for Library and Information Science Education, Boston, Massachusetts.
- Hsieh-Yee, I. (2005, January). *Electronic Resource Management: Practice, Employer Expectations, & CE Interests*. Paper presented at Technical Services Education SIG, 2005 Conference of the Association for Library and Information Science Education, Boston, Massachusetts.
- Hsieh-Yee, I. (2004, October). *Library Professionals for the Digital Age: Competencies & Preparation*. Paper presented at Bibliographic Access Management Team meeting, Library of Congress, Washington, D.C.
- Hsieh-Yee, I. (2004, January). *Cataloging and metadata expertise for the digital era*. Presented at Preparing 21st Century Cataloging and Metadata Professionals: A Workshop for Educators and Trainers, San Diego and sponsored by ALCTS, ALISE, LC, and OCLC.
- Hsieh-Yee, I. (2004, January). *Educating catalogers for the digital era*. Paper presented at the Technical Services SIG, 2004 Conference of the Association for Library and Information Science Education, San Diego.
- Hsieh-Yee, I. (2003, July). *Cataloging Education for the 21st Century*. A presentation at the Library of Congress, Washington, D.C.
- Hsieh-Yee, I. (2002, January) *Metadata Education and Research Priorities: A Delphi Study of Metadata Experts*. Presentation at the 2002 Conference of the Association for Library and Information Science Education, New Orleans.
- Hsieh-Yee, I. (2001, November). *A Delphi Study of Metadata: Preliminary Findings*. Poster session at the 2001 Annual Meeting of the American Society for Information Science & Technology, Washington, D.C.
- Hsieh-Yee, I. (2001, June). *Resources on Asian American Children: Analysis of Retrieval by Search Engines and WorldCat*. Presentation at the National Conference on Asian Pacific American Librarians, San Francisco.
- Hsieh-Yee, I. (2001, January). *Delphi Study on Metadata: Project Design*. Presentation at Research Awards Session, Association for Library & Information Science Education, Washington, D.C.
- Hsieh-Yee, I. (2000, May). *Web Search Behavior Research: Progress and Implications*. Presentation at the Symposium on Evaluating Library and Information Science Research, University of Wisconsin-Madison, Madison, Wisconsin.

- Hsieh-Yee, I. (2000, March). *ERIC User Services: Evaluation in a Decentralized Environment*. Presentation at the National ERIC Joint Directors/Technical Meeting, Arlington, Virginia.
- Hsieh-Yee, I. (2000, January). *Enhancing Learning with Web Technology*. Presentation at Faculty Conversations, Catholic University of America, Washington, D.C.
- Hsieh-Yee, I. (2000, January). *From Surrogates to Objects: CUA's Approaches to Organizing Electronic Resources*. Paper presentation at the Annual Conference of the Association for Library and Information Science Education, San Antonio, Texas.
- Yee, P., and Hsieh-Yee, I. (1997, November). *Individual Differences in Search Behavior on the WWW*. A poster session presented at the 38th Annual Meeting of the Psychonomic Society, Philadelphia, Pennsylvania.
- Hsieh-Yee, I. (1997, April). *Research + Marketing + Preparation = Job!* Presented at the "Workshop on Resume and Interview Techniques," Special Libraries Association, Student Chapter, Catholic University of America, Washington, D.C.
- Hsieh-Yee, I. (1997, February). *Creating CyberCatalogers: Education and Training*. Presentation at ALA's Midwinter Meeting, Washington, D.C.
- Hsieh-Yee, I. (1997, February). *Search Tactics of Web Users in Searching for Texts, Graphics, Known Items and Subjects: A Search Simulation Study*. Presented at the Conference of the Association for Library and Information Science Education, Washington, D.C.
- Hsieh-Yee, Ingrid. "Beginning Your Special Library/Information Center Career." Presented at SLA's "Career Day," Jan. 11, 1997, Catholic University of America.
- Hsieh-Yee, I. (1996, September). *The Roles of Library and Information Scientists in Managing Electronic Information*. Presentation at Hamilton College, Clinton, New York.
- Hsieh-Yee, I. (1996, May). *The Future of Cataloging as a Profession*. Presented at "The Cataloging Forum, Library of Congress, Washington, D.C.
- Hsieh-Yee, I. (1994, October). *The Impact of the Internet on OPACs*. Presented at the Third Workshop on User Interfaces for OPACs, Library of Congress, Washington, D.C.

Reports

- Hsieh-Yee, I., with Knowledge Management Competencies and Performance Action Group of the Federal Knowledge Management Initiative. "From Knowledge Management Competencies to Improved Organizational Performance." April 9, 2009.

Hsieh-Yee, I., with Knowledge Practices Action Group of the Federal Knowledge Management Initiative. "KM Practice in Government Agencies: Findings and Recommendations." April 9, 2009.

Hsieh-Yee, I. "Delphi Study on Metadata." 2001. Three quarterly reports submitted to the Association for Library and Information Science Education.

Hsieh-Yee, I. "College Students' Information Channels: Patterns of Use and Possible Factors in Channel Selection." 1995. Submitted to the Catholic University of America.

Hsieh-Yee, I. "The Information-Seeking Patterns of Scholars and Their Use of an Online Information System." 1994. Submitted to the Council on Library Resources.

Book Reviews

Review of *The Measurement and Evaluation of Library Services*, by Sharon L. Baker and F. Wilfrid Lancaster. *Information Processing and Management* 30 (1994): 450-52.

Review of *Subject Access to Films and Videos*, by Sheila S. Intner and William E. Studwell; and *Cataloging Unpublished Nonprint Materials*, by Verna Urbanski with Bao Chu Chang and Bernard L. Karon. *Information Processing and Management* 30 (1994): 449-50.

Review of *Automated Information Retrieval in Libraries: A Management Handbook*, by Vicki Anders. *Journal of Library and Information Science* 19 (1993): 98-100.

Review of *Full Text Databases*, by Carol Tenopir and Jung Soon Ro. *Information Processing and Management* 28 (1992): 667-68.

Review of *Descriptive Cataloging for the AACR2R And USMARC: A How-to-Do It Workbook*, by Larry Millsap and Terry Ellen Ferl. *Information Processing and Management* 28 (1992): 809-11.

Review of *MARC Manual: Understanding and Using MARC Records*, by Deborah J. Byrne. *Information Processing and Management* 28 (1992): 537-38.

Service

Professional Associations and Societies

- Library of Congress. RDA Training Program for the Profession. Co-authored with Tim Carlton. 2013-2014.
- 2014 Digital Preservation Outreach & Education Survey. Contributed to the design of the survey, 2014.
- National Digital Stewardship Alliance. Outreach Committee. 2011-2014.
- National Digital Stewardship Residency Program. Advisory Group, 2012-2013.
- FEDLINK Health Information Technology Advisory Council, 2011-2015.

- 2012 Joint Conference on Digital Libraries. Program Planning Committee, Pre-Conference Proposals Review Committee, 2012
- Catholic Research Resources Alliance. Five-Year Strategic Plan Task Force, 2011-2012
- Institute of Museum and Library Services. Grant reviewer. 2004, 2005, 2010.
- Association for Library and Information Science Education.
 - * ALISE Bodan Wynar Research Paper Award Committee, 2015, 2016, 2017
 - * ALISE Eugene Garfield Dissertation Award Competition, Jury, 2013, 2014
 - * ALISE Research Grant Competition Committee. Chair, 2012
 - * Pratt-Severn Faculty Innovation Award. Chair, 2009, 2010
 - * ALISE Doctoral Poster Jury, 2012
 - * “Information Organization Competencies for the 21st Century” Discussion session leader. 2009 Conference of the Association for Library and Information Science Education.
 - * Assisted Technical Services SIG Convener in organizing a program, “Building a Metadata Education and Research Community through MERIC (Metadata Education and Research Information Commons): Demo and Stakeholder Input” for the 2008 ALISE conference.
 - * Association for Library Collections and Technical Services/Association for Library and Information Science Education (ALCTS/ALISE) Metadata Education and Research Information Center (MERIC) Advisory Board, Co-Chair (with Sherry Vellucci), 2005-2007. Chair, 2008-2009 (leading the effort to build MERIC, a repository and collaborative space for metadata educators, practitioners, and researchers)
 - * Technical Services SIG, Convener, 2004-2005. Organized a program on “Electronic Resources Management: Current Practices, Employer Expectations, and Teaching Strategies” for the 2005 conference in Boston, Massachusetts.
 - * Technical Services SIG, Convener, 2003-2004. Organized a program on “Organizing Information with Metadata: Desired Competencies and Teaching Innovations” for the 2004 conference.
 - * Technical Services SIG, Convener, 1999-2000. Organized a program on "Teaching the Organization of Electronic Resources" for the 2000 conference.
 - * Curriculum SIG, Co-convener (with Sibyl Moses), 1996-97. Organized a program on “Government Information Policy” for the 1997 conference.
- American Society for Information Science & Technology.
 - * Reviewer, Conference program panel submissions and poster submissions, 2005, 2006, 2007, 2009, 2011, 2012, 2013, 2014, 2015, 2016, 2017
 - * Nomination Committee, 2009-2011
 - * Information Science Education Special Interest Group. American Society for Information Science and Technology. Chair-Elect, 2007-2008. Chair 2008-2009.
 - * Committee on Information Science Education. 1999-2006.
 - * Committee on Information Science Education. Organizing Committee for an orientation program for students at ASIS annual meetings, 1999-2001
 - * Committee on Information Science Education. Sub-committee on Student Welfare (focusing on issues related to master's education), 1998-2001

- * SIG ED. Organizing Committee for the "Seminar on Research and Career Development" for junior researchers. 1995-96 (chair), 1997-2001
- * ISI Doctoral Dissertation Proposal Scholarship Jury, 1997; 2001, 2002
- * Pratt-Severn Best Student Research Paper Award Jury. Chair. 1997
- * 1998 Midyear Meeting (referee of contributed papers), 1997
- * Organizer and moderator of the ASIS Doctoral Forum and the Doctoral Research Seminar 1994-1995
- * SIG Human Computer Interaction. Chair-Elect, Chair, 1993-1995
- * Doctoral Forum Award Jury, 1995
- * Best Student Paper Award Jury, 1995
- American Library Association.
 - * Committee on Accreditation, External Review Panelist, 2009- (site visiting team 2013-2014; site visiting team 2016-2017)
 - * Association for Library Collections and Technical Services Task Force on Competencies and Education for a Career in Cataloging, member, 2008-2009
 - * Facilitator for "What They Don't Teach in Library School: Competencies, Education and Employer Expectations for a Career in Cataloging," an Association for Library Collections and Technical Services Preconference, June 22, 2007 in Washington, D.C. Also a local liaison for bringing this program to the Catholic University of America.
 - * Facilitator for a discussion on "Effect of Electronic Resources on Technical Services" at ALA's Midwinter Meeting held in Feb. 1997 in Washington, D.C.
 - * International Relations Committee, Subcommittee Task Force for IFLA and China, 1994-1997
- Virginia Association of School Librarians. Scholarships and Awards Committee. 2010-2012
- Federal Knowledge Management Initiative, Knowledge Management Practices Action Group. Member. 2009 (leading the effort to build a knowledge management repository)
- Federal Knowledge Management Initiative, Knowledge Management Competencies & Learning Action Group. Member. 2009 (developing an action plan for helping government knowledge workers and government agencies to develop knowledge management competencies)
- National Center for Education Statistics. Technical Review Panel. 2008.
- External evaluator for a case of promotion to full professorship. University of Tennessee. 2008.
- National Information Standards Organization (NISO). Advisory Board, Revision of "IMLS Framework of Guidance for Building Good Digital Collections," 2004, 2007.
- Library of Congress, Bibliographic Control of Web Resources: A Library of Congress Action Plan. Principal Investigator of Action Item 5.1, focusing on cataloging and metadata education for students and new librarians, 2002-2003. (worked with the Association for Library Collections and Technical Services, Education Task Force)

- Chinese American Librarians Association
 - * Chinese American Librarians Association Outstanding Library Leadership Award in Memory of Dr. Margaret Chang Fung, Award Committee, 2016-2017
 - * Achievement Award Jury, 2000-2001
 - * CALA Goal 2000 Task Force, 1997
 - * Scholarship Committee, 1995, 1996-1997 (chair)
 - * Board of Directors, 1994-1997
 - * Publication Committee, 1993-1995
 - * International Relations Committee, 1993-1996
- SailorSM Assessment Advisory Group (An impact study of Sailor, Maryland's Public Information Network), 1995
- Editorial boards
 - Journal of Library & Information Science. Editorial Board, 2012-
 - Chinese American Librarians Association, *Occasional Papers Series*. Editorial Board, 2009-2016.
 - Library Quarterly*. Editorial Board, 2003-2008
 - Bulletin of the Medical Library Association*, 1994-97
 - Newsletter editor for the Chinese American Librarians Association, 1989-92
- Referee for the following journals
 - Information Processing and Management*
 - Journal of Digital Information*
 - Journal of Education for Library and Information Science*
 - Journal of Information Science*
 - Journal of Library & Information Science*
 - Journal of Library Metadata*
 - Journal of the American Society for Information Science & Technology*
 - Library and Information Science Research*
 - Library Quarterly*
- Expert reviewer, “Digital Library” course, Evaluation module, University of North Carolina, Chapel Hill, 2007-2008.
- Expert reviewer, “Information Organization” course, University of Michigan, Ann Arbor. 2007.

Catholic University of America

- Academic Senate Committee on Appointments and Promotions, 2005-2008
- Academic Senate Committee on Committees and Rules, 2009-2012, 2019-2022
- Academic Senate representative (for School of Arts & Sciences), 2017-2020

- Academic Senate representative (for School of Library & Information Science) 2003-2012
- Academic Senate Library Committee, Interim Chair (2007), Member, 2008-2012
- Academic Senate Committee on Computing, 1995-2003
- President's Administrative Council, 2010-2012
- Deans' Council, 2010-2012
- Academic Leadership Group, 2010-2012
- Graduate Board, 2010-2012
- School of Arts & Sciences, Committee on Appointments and Promotions, 2015-2019
- School of Arts & Sciences, Academic Council, 2015-2016.
- School of Arts & Sciences, Ordinary Professor Group, 2013-
- Doctoral Dissertation Defense Committee, Chair, Dept. of Psychology, 2016, 2017, 2018, 2019, 2020, 2021, 2022
- Doctoral Dissertation Defense Committee, Chair, Dept. of Education, 2014, 2015, 2017, 2018, 2019
- Doctoral Dissertation Defense Committee, Chair, National Catholic School of Social Services, 2019, 2020, 2021
- Doctoral Dissertation Defense Committee, Chair, School of Nursing, 2006, 2008
- CUA Scholarly Communication Project Team, Member (2007), Chair, 2008-2009
- Dean Search Committee, 1992-1994, 1998-1999, 2002-2003, 2006-2007
- Fulbright Review Panel, 2006
- CUA Service Learning Advisory Board, 2001-2002
- CUA Faculty Conversations on Enhancing Teaching and Learning through Technology, Planning Group, 1999-2001
- CUA Initiative on Technology and Teaching, 1998-2001

Dept. of Library and Information Science

- Admissions Committee, 2007-2009, Chair 2010-2012, Member 2013-2015, Member 2018-2022
- Appointments and Promotions Committee, 1991-2022
- Student advisement, 1990-2022
- 4+1 Bachelor to Master program, Program Advisor, 2014-2022
- LIS Minor, Advisor, 2017-2022
- Curriculum Committee, 2020-21 (Member), 2021-2022 (Chair)
- Scholarship & Awards Committee, 2016-2018, 2020-21
- Comprehensive examination editor, 2020-21
- Symposium, Colloquium, Lecture Series Committee, May 2018-May 2020 (Chair)
- Symposium and Colloquium Committee, fall 2016-May 2018
- Bridging the Spectrum Symposium session moderator, 2018, 2019, 2020, 2021

- Organization of Information Core Course Outcomes Assessment, 2019, 2020
- The Information Professions in Society Core Course Outcomes Assessment, 2019, 2020
- Community Services Librarianship, Course of Study Review and Revision, 2020
- Organization of Information, Course of Study Review and Revision, 2019
- Organization of Information, Competency and Course Offerings Review, 2018-2019
- Accreditation presentation, Chair, June 2015-August 2016
- Interim Co-Chair, June 2015-August 2016.
- Blended/OWL Learning Committee, spring 2016-2018
- Technology Committee, fall 2016-2017
- Comprehensive examination editor, 2016-2017, reader (every year since 1990)
- LIS Advisory Board, 2015-2016 (chair); fall 2016- May 2018 (member)
- Committee on Planning and Assessment, 2015-2016 (chair)
- Senior Faculty Committee, 2014-2016.
- Accreditation Steering Committee, 2014-2016 (Chair, 2015-2016)
- Accreditation Students Standard Committee, co-chair, 2014-2016
- Accreditation Mission, Goals, and Objectives Standard Committee, co-chair, 2014-2016
- Accreditation Curriculum Standard, member 2014-2-16
- Accreditation Administration and Finance Standard, member 2014-2016
- Cultural Heritage Information Management Project (IMLS-funded), Co-PI, 2012-2015
- Cultural Heritage Information Management Forum (scheduled for June 2015), Co-Organizer, 2013-2015
- Health Information Technology Interim Review Committee, 2015 (chair)
- Health Sciences Librarianship Advisory Group, 2015- (chair)
- Comprehensive examination editors, 2013-2014, 2016-2017
- National Digital Stewardship Alliance liaison, 2011-2014
- Advisory Board, Chair 2010-2012
- Academic Honesty Committee, Chair, 2008-2012
- Blended Learning Committee, 2010-2012
- Colloquium Committee, 2010-2012
- Comprehensive Examination Administration, 2010-2012
- Cultural Heritage Information Management Advisory Committee, 2010-2012 (chair), 2013-
- Curriculum Committee, 1991-2003, 2007-2009, Chair 2010-2012, member 2013-
- Curriculum Subcommittee on Comprehensive Examination, Chair 2009-2012
- Health Information Technology Advisory Board, Chair 2010-2012. Member 2013-
- Health Sciences Advisory Committee, 2009, Chair 2010-2012. Member 2013-
- HIT Expert Forum, Chair 2012. Member 2013-2015
- Health Information Technology Student Group Advisor, 2011-2012
- State Council for Higher Education of Virginia, SLIS Representative, 2010-2012
- Symposium Planning Committee, 2010-2012
- Website Management Team, Chair, 2010-2012
- Urban School Librarianship Project (IMLS-Funded), PI, 2007-2011 (chair, 2010-11)

- Failing Grades Committee, 1995-1997 (chair), 2000-2001 (chair), 2004-2005 (chair), 2007 (chair)-2011
- Faculty Search Committee, 1994-1998, 2002-2004, 2006 (chair), Fall 2007-2009, Chair fall 2009-2012
- Recruitment Committee, Chair 2010-2012
- Strategic Planning Committee, Chair 2010-2012
- Technology Committee, 2010-2012
- Accreditation Advisory Committee, 2007-2009
- Accreditation Coordinating Committee, 2007-2009
- Accreditation Steering Committee, 2007-2009
- SLIS Advisory Group, 2007-2009
- Accreditation Curriculum Standard Committee, Co-chair, 2007-2009
- Accreditation Faculty Standard Committee, Co-chair, 2007-2009
- LSC 551 Information Organization Review Team, Co-chair, 2008-2009, 2015-2016.
- Curriculum Subcommittee on Portfolios, 2009
- LSC 555 Information Systems in Libraries and Information Centers Review Team, contributor, 2008-2009
- Redesign of LSC 730 Use and Users of Libraries and Information. 2009-
- Development of a metadata institute that was taught as LSC 715 Organization of Internet Resources in 2008. The institute is being revised and will be offered in 2010 under a new course title.
- Development of lesson plans, assignments, and evaluation rubrics for LSC 606, Cataloging and Classification, for the School's NCATE accreditation. 2008
- Howard and Mathilde Rovelstad Scholarship Committee, Chair, 2004-2007
- Assistant Dean Search Committee, Chair, Fall 2007
- Liaison to the Association for Library Collections and Technical Services to bring its preconference program, Cataloging Education and Employer Expectations, to CUA during the 2007 American Library Association Annual Meeting in Washington, D.C.
- Organizer of the colloquium presentation and reception for Tamar Sadeh of Ex Libris on PRIMO June 2007
- Practicum review and design (work with potential supervisors, such as the American Indian Museum internship description revision) 2006-
- Comprehensive examinations (edits, proctoring, and grading), 1990-
- SLIS Web site redesign: Comments and suggestions. Fall 2007
- Conducted surveys of current students and alumni in preparation for the 2005 re-accreditation, 2004-2005
- Technology Committee, 1992-1999 (chair, 1996-1998), 2002-2003 (member)
- Colloquia Committee 1997-1999, 2002-2003.
- Advisor of the CUA Student Chapter of the American Society for Information Science and Technology, 2002-2003
- Visiting Professor Search Committee, 1999, 2000, 2001
- Leader, Participation in the CORC experiment, 1999-2000
- Advisor of the Special Libraries Association Student Chapter, 1993-1999; the group was recognized for outstanding leadership by SLA in 1999.

- COA planning Committee, Task Force on Electronic Presentation of SLIS Reports (team leader) 1997-1998
- COA Planning Committee, Subcommittee on Technology 1996-1998
- NLM practicum coordinator, 1997-1998
- Computer Literacy Workshops: Assisted with the development and evaluation of the workshops, 1996-1998
- Leader, Participation in the InterCat project, 1995-1997

FULL TEXT LINKS

Science Translational Medicine

Sci Transl Med. 2010 Feb 24;2(20):20ra14. doi: 10.1126/scitranslmed.3000702.

Development of personalized tumor biomarkers using massively parallel sequencing

Rebecca J Leary¹, Isaac Kinde, Frank Diehl, Kerstin Schmidt, Chris Clouser, Cisilya Duncan, Alena Antipova, Clarence Lee, Kevin McKernan, Francisco M De La Vega, Kenneth W Kinzler, Bert Vogelstein, Luis A Diaz Jr, Victor E Velculescu

Affiliations

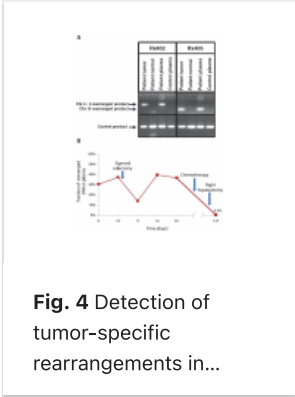
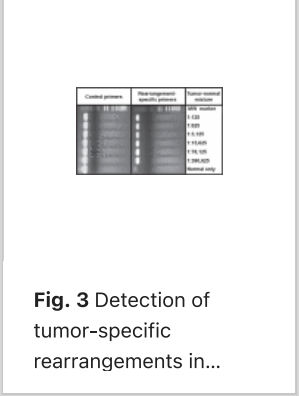
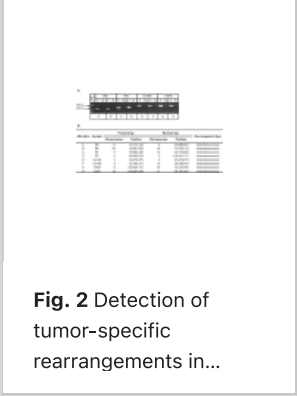
PMID: 20371490 PMCID: [PMC2858564](#) DOI: [10.1126/scitranslmed.3000702](#)

[Free PMC article](#)

Abstract

Clinical management of human cancer is dependent on the accurate monitoring of residual and recurrent tumors. The evaluation of patient-specific translocations in leukemias and lymphomas has revolutionized diagnostics for these diseases. We have developed a method, called personalized analysis of rearranged ends (PARE), which can identify translocations in solid tumors. Analysis of four colorectal and two breast cancers with massively parallel sequencing revealed an average of nine rearranged sequences (range, 4 to 15) per tumor. Polymerase chain reaction with primers spanning the breakpoints was able to detect mutant DNA molecules present at levels lower than 0.001% and readily identified mutated circulating DNA in patient plasma samples. This approach provides an exquisitely sensitive and broadly applicable approach for the development of personalized biomarkers to enhance the clinical management of cancer patients.

Figures



Comment in

Cancer sequencing gets a little more personal.

Prokunina-Olsson L, Chanock SJ.

Sci Transl Med. 2010 Feb 24;2(20):20ps8. doi: 10.1126/scitranslmed.3000929.

PMID: 20371489

Related information

MedGen

PubChem Compound (MeSH Keyword)

LinkOut – more resources

Full Text Sources

Atypon

Europe PubMed Central

PubMed Central

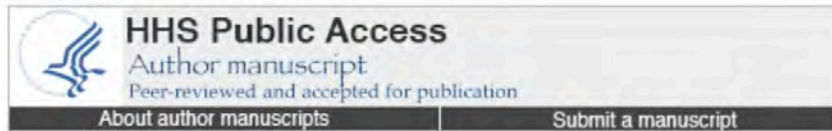
Other Literature Sources

The Lens - Patent Citations

Medical

ClinicalTrials.gov

[Journal List](#) > [HHS Author Manuscripts](#) > PMC2858564



[Sci Transl Med](#). Author manuscript; available in PMC 2010 Aug 24.

Published in final edited form as:

[Sci Transl Med](#). 2010 Feb 24; 2(20): 20ra14.

doi: [10.1126/scitranslmed.3000702](#)

PMCID: PMC2858564

NIHMSID: NIHMS192215

PMID: [20371490](#)

Development of Personalized Tumor Biomarkers Using Massively Parallel Sequencing

[Rebecca J. Leary](#),¹ [Isaac Kinde](#),¹ [Frank Diehl](#),¹ [Kerstin Schmidt](#),¹ [Chris Clouser](#),² [Cisilya Duncan](#),² [Alena Antipova](#),² [Clarence Lee](#),² [Kevin McKernan](#),² [Francisco M. De La Vega](#),³ [Kenneth W. Kinzler](#),¹ [Bert Vogelstein](#),¹ [Luis A. Diaz, Jr.](#),¹ and [Victor E. Velculescu](#)^{1,*}

▶ [Author information](#) ▶ [Copyright and License information](#) [Disclaimer](#)

The publisher's final edited version of this article is available at [Sci Transl Med](#)

See commentary "[Cancer sequencing gets a little more personal](#)." in *Sci Transl Med*, volume 2 on page 20ps8.

OTHER FORMATS

[PubReader](#) | [PDF \(1.0M\)](#)

ACTIONS

“ Cite

☆ Favorites

SHARE



RESOURCES

Similar articles +

Cited by other articles +

Links to NCBI Databases +

Associated Data

PMID- 20371490
OWN - NLM
STAT- MEDLINE
DCOM- 20100615
LR - 20220321
IS - 1946-6242 (Electronic)
IS - 1946-6234 (Print)
IS - 1946-6234 (Linking)
VI - 2
IP - 20
DP - 2010 Feb 24
TI - Development of personalized tumor biomarkers using massively parallel sequencing
PG - 20ra14
LID - 10.1126/scitranslmed.3000702 [doi]
AB - Clinical management of human cancer is dependent on the accurate monitoring of residual and recurrent tumors. The evaluation of patient-specific translocations in leukemias and lymphomas has revolutionized diagnostics for these diseases. We have developed a method, called personalized analysis of rearranged ends (PARE), which can identify translocations in solid tumors. Analysis of four colorectal and two breast cancers with massively parallel sequencing revealed an average of nine rearranged sequences (range, 4 to 15) per tumor. Polymerase chain reaction with primers spanning the breakpoints was able to detect mutant DNA molecules present at levels lower than 0.001% and readily identified mutated circulating tumor DNA in patient plasma samples. This approach provides an exquisitely sensitive and broadly applicable approach for the development of personalized biomarkers to enhance the clinical management of cancer patients.
FAU - Leary, Rebecca J
AU - Leary RJ
AD - Ludwig Center for Cancer Genetics and Therapeutics and Howard Hughes Medical Institute, Johns Hopkins Kimmel Cancer Center, Baltimore, MD 21231, USA
FAU - Kinde, Isaac
AU - Kinde I
FAU - Diehl, Frank
AU - Diehl F
FAU - Schmidt, Kerstin
AU - Schmidt K
FAU - Clouser, Chris
AU - Clouser C
FAU - Duncan, Cisilya
AU - Duncan C
FAU - Antipova, Alena
AU - Antipova A
FAU - Lee, Clarence
AU - Lee C
FAU - McKernan, Kevin
AU - McKernan K
FAU - De La Vega, Francisco M
AU - De La Vega FM
FAU - Kinzler, Kenneth W
AU - Kinzler KW
FAU - Vogelstein, Bert
AU - Vogelstein B
FAU - Diaz, Luis A Jr

AU - Diaz LA Jr
FAU - Velculescu, Victor E
AU - Velculescu VE
LA - eng
GR - R37 CA043460/CA/NCI NIH HHS/United States
GR - P50 CA062924/CA/NCI NIH HHS/United States
GR - HHMI/Howard Hughes Medical Institute/United States
GR - R01 CA121113/CA/NCI NIH HHS/United States
GR - R01 CA121113-05/CA/NCI NIH HHS/United States
GR - R37 CA057345/CA/NCI NIH HHS/United States
GR - R37 CA057345-18/CA/NCI NIH HHS/United States
GR - R37 CA043460-27/CA/NCI NIH HHS/United States
GR - R01 CA057345/CA/NCI NIH HHS/United States
PT - Journal Article
PT - Research Support, Non-U.S. Gov't
TA - Sci Transl Med
JT - Science translational medicine
JID - 101505086
RN - 0 (Biomarkers, Tumor)
RN - 9007-49-2 (DNA)
SB - IM
CIN - Sci Transl Med. 2010 Feb 24;2(20):20ps8. PMID: 20371489
MH - Base Sequence
MH - Biomarkers, Tumor/blood/*genetics
MH - DNA/blood/genetics
MH - *Gene Rearrangement
MH - Humans
MH - Molecular Sequence Data
MH - Neoplasms/*genetics
MH - Polymerase Chain Reaction/*methods
MH - *Precision Medicine/instrumentation/methods
MH - Sequence Analysis, DNA/*methods
MH - *Translocation, Genetic
PMC - PMC2858564
MID - NIHMS192215
EDAT- 2010/04/08 06:00
MHDA- 2010/06/16 06:00
CRDT- 2010/04/08 06:00
PHST- 2010/04/08 06:00 [entrez]
PHST- 2010/04/08 06:00 [pubmed]
PHST- 2010/06/16 06:00 [medline]
AID - 2/20/20ra14 [pii]
AID - 10.1126/scitranslmed.3000702 [doi]
PST - ppublish
SO - Sci Transl Med. 2010 Feb 24;2(20):20ra14. doi: 10.1126/scitranslmed.3000

[Home](#) > [Science Journals](#) > [Science Translational Medicine Home](#)

Search the Journal

Vol#

Pg#

[Advanced >](#)

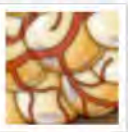
Contributing to Sci TM

[Submit a Manuscript](#)



[Table of Contents](#)

HIGHLIGHTS



PERSPECTIVE

Adipose Tissue Macrophages: A Piece of the PAI of Metabolic Syndrome

Adipose tissue macrophages may link cardiovascular disease and diabetes via the dynamic production of plasminogen activator inhibitor-1.

Read the related Research Article:
[Adipocyte-Derived Factors Potentiate Nutrient-Induced Production of Plasminogen Activator Inhibitor-1 by Macrophages](#)



RESEARCH

Development of Personalized Tumor Biomarkers Using Massively Parallel Sequencing

[To Advertise](#) [Find Products](#)

AAAS.ORG | FEEDBACK | HELP | LIBRARIANS

Science Translational Medicine

Enter Search Term

ADVANCED

GUEST | ALERTS | ACCESS RIGHTS | MY ACCOUNT | SIGN IN |

[Sci TM Home](#) | [Current Issue](#) | [Issue Archive](#) | [Multimedia](#) | [Community Resources](#) | [My Sci TM](#) | [About Sci TM](#)
[Home](#) > [Science Journals](#) > [Science Translational Medicine Home](#)

Search the Journal

Vol#

Pg# [GO >](#)

Enter Keyword

Full text ☐

[GO >](#)

[Advanced >](#)

Contributing to *Sci TM*[Submit a Manuscript](#)About *Sci TM*

[Mission](#)

[Advisory Board](#)

[Associate Advisors](#)

[Staff](#)

[Editorial Policies](#)

[Contact Us](#)

Information for

[Authors](#)

[Referees](#)

[Librarians/Institutions](#)

[Advertisers](#)

Reader Services

[Join AAAS/Subscribe](#)

[Free Registration](#)

[Free Two Week Trial](#)

[Email Alerts](#)

[RSS Feeds](#)

[Reprints and Permissions](#)

My Sci TM

[My Folders](#)

[My Alerts](#)

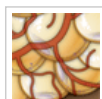
[My Saved Searches](#)

[Sign In](#)



about the cover

HIGHLIGHTS



PERSPECTIVE

Adipose Tissue Macrophages: A Piece of the PAI of Metabolic Syndrome

Adipose tissue macrophages may link cardiovascular disease and diabetes via the dynamic production of plasminogen activator inhibitor-1.

Read the related Research Article:
[Adipocyte-Derived Factors Potentiate Nutrient-Induced Production of Plasminogen Activator Inhibitor-1 by Macrophages](#)



RESEARCH

Development of Personalized Tumor Biomarkers Using Massively Parallel Sequencing

Rapid detection of specific aberrant rearrangements in tumors from individuals leaps towards personalized oncology.

Read the related Perspective:
[Cancer Sequencing Gets a Little More Personal](#)

Also of Interest:

SCIENCE POLICY FORUM:
[NIH GUIDELINES FOR STEM CELL RESEARCH AND GAMETE DONORS](#)

Coming Next Week:

Seeing is believing: hope for retinal degeneration.

Watch our video
introducing the new
journal.

Table of Contents

[Perspectives](#)

[Research](#)

[Editors' Choice](#)

FROM *SCIENCE* CAREERS

The Clinical and Translational Science Network (CTSciNet) provides articles and perspectives about training, career paths, and career-related issues in clinical and translational science, and an online community open to all who are interested in careers in clinical and translational research. [Find out more](#)

Join the CTSciNet Community

FEATURED ARTICLE
[Informatics Careers Take Shape in Translational and Clinical Research](#)
 By Brian Vastag

FEATURED ARTICLE
[Perspective: Three Crucial Questions When Applying to M.D.-Ph.D. Programs](#)
 By Skip Brass

© 2010 American Association for the Advancement of Science. All Rights Reserved.
AAAS is a partner of [HINARI](#), [AGORA](#), [OARE](#), [eIFL](#), [PatientInform](#), [CrossRef](#), and [COUNTER](#).

Home > [Science Journals](#) > [Science Translational Medicine Home](#) > [24 February 2010](#) > [Leary et al., 2:\(20\): 20ra14](#)

- Article Views
- [Abstract](#)
- [Full Text](#)
- [Full Text \(PDF\)](#)
- [Supplementary Material](#)
- Article Tools
- [Save to My Folders](#)
- [Download Citation](#)
- [Alert Me When Article](#)

Sci Transl Med 24 February 2010:
Vol. 2, Issue 20, p. 20ra14
Sci. Transl. Med. DOI: 10.1126/scitranslmed.3000702

[< Prev](#) | [Table of Contents](#) | [Next >](#) [To Advertise](#) [Find Products](#)

RESEARCH ARTICLE

CANCER GENETICS

Development of Personalized Tumor Biomarkers Using Massively Parallel Sequencing

Rebecca J. Leary¹, Isaac Kinde¹, Frank Diehl¹, Kerstin Schmidt¹, Chris Clouser², Cisilya Duncan², Alena Antipova², Clarence Lee², Kevin McKernan², Francisco M. De La Vega³, Kenneth W. Kinzler¹, Bert Vogelstein¹, Luis A. Diaz Jr.¹ and Victor E. Velculescu^{1,*}

+ Author Affiliations

AAAS.ORG | FEEDBACK | HELP | LIBRARIANS

Science Translational Medicine ▼

Enter Search Term

ADVANCED

[Sci TM Home](#) [Current Issue](#) [Rapid Publication](#) [Issue Archive](#) [Multimedia](#) [Sci TM Collections](#) [My Sci TM](#) [About Sci TM](#)[Home](#) > [Science Journals](#) > [Science Translational Medicine Home](#) > [24 February 2010](#) > [Leary et al., 2:\(20\): 20ra14](#)**Science Translational Medicine**[Prev](#) | [Table of Contents](#) | [Next](#)

stm.sciencemag.org

Sci Transl Med 24 February 2010:

Vol. 2, Issue 20, p. 20ra14

Sci. Transl. Med. DOI: 10.1126/scitranslmed.3000702

• **RESEARCH ARTICLE****CANCER GENETICS****Development of Personalized Tumor Biomarkers Using Massively Parallel Sequencing**

Rebecca J. Leary¹, Isaac Kinde¹, Frank Diehl¹, Kerstin Schmidt¹, Chris Clouser², Cisilya Duncan², Alena Antipova², Clarence Lee², Kevin McKernan², Francisco M. De La Vega³, Kenneth W. Kinzler¹, Bert Vogelstein¹, Luis A. Diaz Jr.¹ and Victor E. Velculescu^{1,*}

[+ Author Affiliations](#)*To whom correspondence should be addressed. E-mail: velculescu@jhmi.edu**ABSTRACT**

Clinical management of human cancer is dependent on the accurate monitoring of residual and recurrent tumors. The evaluation of patient-specific translocations in leukemias and lymphomas has revolutionized diagnostics for these diseases. We have developed a method, called personalized analysis of rearranged ends (PARE), which can identify translocations in solid tumors. Analysis of four colorectal and two breast cancers with massively parallel sequencing revealed an average of nine rearranged sequences (range, 4 to 15) per tumor. Polymerase chain reaction with primers spanning the breakpoints was able to detect mutant DNA molecules present at levels lower than 0.001% and readily identified mutated circulating DNA in patient plasma samples. This approach provides an exquisitely sensitive and broadly applicable approach for the development of personalized biomarkers to enhance the clinical management of cancer patients.

Received December 7, 2009. Accepted February 4, 2010.

Copyright © 2010, American Association for the Advancement of Science

Citation: R. J. Leary, I. Kinde, F. Diehl, K. Schmidt, C. Clouser, C. Duncan, A. Antipova, C. Lee, K. McKernan, F. M. Vega, K. W. Kinzler, B. Vogelstein, L. A. Diaz, V. E. Velculescu, Development of Personalized Tumor Biomarkers Using Massively Parallel Sequencing. *Sci. Transl. Med.* 2, 20ra14 (2010).

[Read the Full Text](#)**THE EDITORS SUGGEST THE FOLLOWING RELATED RESOURCES ON SCIENCE SITES**[BACK TO TOP](#)**In Science Translational Medicine****PERSPECTIVE:****CANCER****Cancer Sequencing Gets a Little More Personal**

Ludmila Prokunina-Olsson and Stephen J. Chanock

Sci Transl Med 24 February 2010 2:20ps8[Abstract](#) [Full Text](#) [Full Text \(PDF\)](#)

Home > Science Journals > Science Translational Medicine Home > 24 February 2010 > Leary et al., 2:(20): 20ra14

Register for Free or Subscribe/Join AAAS to View Full Text.

The content you requested requires [free registration](#) or a [subscription](#) to this site. If you already have a user name and password, please sign in below.

CANCER GENETICS

Development of Personalized Tumor Biomarkers Using Massively Parallel Sequencing

Rebecca J. Leary, Isaac Kinde, Frank Diehl, Kerstin Schmidt, Chris Clouser, Cisilya Duncan, Alena Antipova, Clarence Lee, Kevin McKernan, Francisco M. De La Vega, Kenneth W. Kinzler, Bert Vogelstein, Luis A. Diaz, Jr., and Victor E. Velculescu

Sci Transl Med 24 February 2010 2:20ra14

[View Free Abstract](#)

NEED HELP?

- [Regain access to a Pay-per-view article](#) >
- [Can't get past this page?](#) >
- [Why don't I have access?](#) >

RECOMMEND A SUBSCRIPTION TO YOUR INSTITUTION.

[RECOMMEND NOW](#) >

Home > Science Journals > Science Translational Medicine Home > 24 February 2010 > Leary et al., 2:(20): 20ra14

Register for Free or Subscribe/Join AAAS to View Full Text.

The content you requested requires [free registration](#) or a [subscription](#) to this site. If you already have a user name and password, please sign in below.

CANCER GENETICS

Development of Personalized Tumor Biomarkers Using Massively Parallel Sequencing

Rebecca J. Leary, Isaac Kinde, Frank Diehl, Kerstin Schmidt, Chris Clouser, Cisilya Duncan, Alena Antipova, Clarence Lee, Kevin McKernan, Francisco M. De La Vega, Kenneth W. Kinzler, Bert Vogelstein, Luis A. Diaz, Jr., and Victor E. Velculescu

Sci Transl Med 24 February 2010 2:20ra14

View Free Abstract

NEED HELP?

[Regain access to a Pay-per-view article >](#)

[Can't get past this page? >](#)

[Why don't I have access? >](#)

RECOMMEND A SUBSCRIPTION TO YOUR INSTITUTION.

RECOMMEND NOW >

How Do I Get Access?

Sign In	Activate	Subscribe/Join	Register (Free)
User Name: Password: GO > ☐ Auto Sign-In: Save my user name & password on this computer. Help with Sign-In Forgotten user name or password?	AAAS members activate your FREE subscription to ALL <i>Science</i> content since 1880, as well as <i>ScienceNOW</i>, the SAGE KE Archive and <i>AAASMember.org</i> . (more info) ACTIVATE >	Receive full access to ALL <i>Science</i> content since 1880, as well as <i>ScienceNOW</i>, the SAGE KE Archive and <i>AAASMember.org</i> SUBSCRIBE >	Full access to <i>Science</i> research papers more than one year old (and published since 1997), the SAGE KE Archive, e-mail alerts, partial access to <i>Science Signaling</i> content, and personalization features for institutional subscribers, are available to you FREE with registration. (more info) REGISTER >

Why Don't I Have Access?

The content you requested requires [free registration](#) or a [subscription](#) to this site. To find out what content you currently have access to – [view your access rights](#).

Appendix 002C

Early Citations to Leary

Source Document

Leary, R. J., Kinde, I., Diehl, F., Schmidt, K., Clouser, C., Duncan, C., Antipova, A., Lee, C., McKernan, K., De La Vega, F. M., Kinzler, K. W., Vogelstein, B., Diaz, L. A., Jr, & Velculescu, V. E. (2010). Development of personalized tumor biomarkers using massively parallel sequencing. *Science translational medicine*, 2(20), 20ra14.

Early Citations from 2010 (in chronological order)

Newman, S., & Edwards, P. A. (2010, March). High-throughput analysis of chromosome translocations and other genome rearrangements in epithelial cancers. *Genome Medicine*, 2(3), article number 19. DOI: 10.1186/gm140.

Robison K. (published online April 28, 2010; print issue date September 2010). Application of second-generation sequencing to cancer genomics. *Briefings in bioinformatics*, 11(5), 524–534.

Mardis E. R. (2010, May 5). Cancer genomics identifies determinants of tumor biology. *Genome biology*, 11(5), 211.

Chmielecki, J., Peifer, M., Jia, P., Socci, N. D., Hutchinson, K., Viale, A., ... & Pao, W. (published online June 29, 2010; print issue date November 2010). Targeted next-generation sequencing of DNA regions proximal to a conserved GXGXXG signaling motif enables systematic discovery of tyrosine kinase fusions in cancer. *Nucleic Acids Research*, 38(20), 6985–6996.

Swanton, C., & Caldas, C. (2010, October). From genomic landscapes to personalized cancer management-is there a roadmap?. *Annals of the New York Academy of Sciences*, 1210, 34–44.

McBride, D. J., Orpana, A. K., Sotiriou, C., Joensuu, H., Stephens, P. J., Mudie, L. J., Hämäläinen, E., Stebbings, L. A., Andersson, L. C., Flanagan, A. M., Durbecq, V., Ignatiadis, M., Kallioniemi, O., Heckman, C. A., Alitalo, K., Edgren, H., Futreal, P. A., Stratton, M. R., & Campbell, P. J. (2010,

November). Use of cancer-specific genomic rearrangements to quantify disease burden in plasma from patients with solid tumors. *Genes, chromosomes & cancer*, 49(11), 1062–1069.

Lo, Y. M., Chan, K. C., Sun, H., Chen, E. Z., Jiang, P., Lun, F. M., Zheng, Y. W., Leung, T. Y., Lau, T. K., Cantor, C. R., & Chiu, R. W. (2010, December). Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Science Translational Medicine*, 2(61), 61ra91.



Any time

Since 2022

Since 2021

Since 2018

Custom range...

Sort by relevance

Sort by date



Create alert

Development of personalized tumor biomarkers using massively parallel sequencing

☐ Search within citing articles

Liquid biopsies come of age: towards implementation of circulating tumour DNA

[J Wan](#), [C Massie](#), [J Garcia-Corbacho](#)... - *Nature Reviews* ..., 2017 - nature.com

Key Points Cell-free DNA (cfDNA) is released predominantly by cell death into the bloodstream, although active secretion may have a role. Since the discovery of fetal cfDNA ...

☆ Save Cite Cited by 1700 Related articles All 8 versions

[PDF] cam.ac.uk
ViewIt@CatholicU

Application of cell-free DNA analysis to cancer treatment

[RB Corcoran](#), [BA Chabner](#) - *New England Journal of Medicine*, 2018 - Mass Medical Soc

Emerging Roles of Cell-free Tumor DNA in Oncology The capacity to detect new cancers, treatment-resistant variants, and tumor heterogeneity by noninvasive technology on the ...

☆ Save Cite Cited by 521 Related articles All 4 versions

ViewIt@CatholicU

Genome-wide cell-free DNA fragmentation in patients with cancer

[S Cristiano](#), [A Leal](#), [J Phallen](#), [J Fiksel](#), [V Adleff](#)... - *Nature*, 2019 - nature.com

Cell-free DNA in the blood provides a non-invasive diagnostic avenue for patients with cancer. However, characteristics of the origins and molecular features of cell-free DNA are ...

☆ Save Cite Cited by 525 Related articles All 18 versions

[HTML] nih.gov
ViewIt@CatholicU

Detection of circulating tumor DNA in early-and late-stage human malignancies

[C Bettegowda](#), [M Sausen](#), [RJ Leary](#), [I Kinde](#)... - *Science translational* ..., 2014 - science.org

The development of noninvasive methods to detect and monitor tumors continues to be a major challenge in oncology. We used digital polymerase chain reaction-based ...

☆ Save Cite Cited by 3873 Related articles All 18 versions

[HTML] nih.gov

Integrating liquid biopsies into the management of cancer

[G Siravegna](#), [S Marsoni](#), [S Siena](#)... - *Nature reviews Clinical* ..., 2017 - nature.com

Key Points Patient selection is central to the success of targeted therapies. Identification of

[PDF] cancer-id.eu
ViewIt@CatholicU

Any time
 Since 2022
 Since 2021
 Since 2018
 Custom range...

2010 — 2010

Search

Sort by relevance

Sort by date

Create alert

Development of personalized tumor biomarkers using massively parallel sequencing

☐ Search within citing articles

Advances in understanding cancer genomes through second-generation sequencing

[M Meyerson](#), [S Gabriel](#), [G Getz](#) - [Nature Reviews Genetics](#), 2010 - [nature.com](#)

Key Points Analyses of cancer genome sequences and structures provide insights for understanding cancer biology, diagnosis and therapy. The application of second-generation ...

☆ Save Cite Cited by 1399 Related articles All 19 versions

[PDF] [umb.edu](#)
[ViewIt@CatholicU](#)

Distant metastasis occurs late during the genetic evolution of pancreatic cancer

[S Yachida](#), [S Jones](#), [I Bozic](#), [T Antal](#), [R Leary](#), [B Fu...](#) - [Nature](#), 2010 - [nature.com](#)

A timeline for pancreatic cancer Christine Iacobuzio-Donahue and colleagues use whole-genome exome sequencing to analyse primary pancreatic cancers and one or more ...

☆ Save Cite Cited by 2641 Related articles All 29 versions

[HTML] [nih.gov](#)
[ViewIt@CatholicU](#)

Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus

[YMD Lo](#), [KCA Chan](#), [H Sun](#), [EZ Chen](#), [P Jiang...](#) - [Science translational ...](#), 2010 - [science.org](#)

Cell-free fetal DNA is present in the plasma of pregnant women. It consists of short DNA fragments among primarily maternally derived DNA fragments. We sequenced a maternal ...

☆ Save Cite Cited by 1190 Related articles All 9 versions

[PDF] [researchgate.net](#)

Next generation DNA sequencing and the future of genomic medicine

[MW Anderson](#), [I Schrijver](#) - [Genes](#), 2010 - [mdpi.com](#)

In the years since the first complete human genome sequence was reported, there has been a rapid development of technologies to facilitate high-throughput sequence analysis of DNA ...

☆ Save Cite Cited by 113 Related articles All 11 versions

[HTML] [mdpi.com](#)
[ViewIt@CatholicU](#)

Use of (Messages if genomic rearrangements to quantify disease burden in plasma from patients with solid tumors

[HTML] [nih.gov](#)

FULL TEXT LINKS



> [Clin Chem](#). 2013 Jan;59(1):211-24. doi: 10.1373/clinchem.2012.196014. Epub 2012 Oct 11.

Cancer genome scanning in plasma: detection of tumor-associated copy number aberrations, single-nucleotide variants, and tumoral heterogeneity by massively parallel sequencing

K C Allen Chan ¹, Peiyong Jiang, Yama W L Zheng, Gary J W Liao, Hao Sun, John Wong, Shing Shun N Siu, Wing C Chan, Stephen L Chan, Anthony T C Chan, Paul B S Lai, Rossa W K Chiu, Y M D Lo

Affiliations

PMID: 23065472 DOI: [10.1373/clinchem.2012.196014](https://doi.org/10.1373/clinchem.2012.196014)

Abstract

Background: Tumor-derived DNA can be found in the plasma of cancer patients. In this study, we explored the use of shotgun massively parallel sequencing (MPS) of plasma DNA from cancer patients to scan a cancer genome noninvasively.

Methods: Four hepatocellular carcinoma patients and a patient with synchronous breast and ovarian cancers were recruited. DNA was extracted from the tumor tissues, and the preoperative and postoperative plasma samples of these patients were analyzed with shotgun MPS.

Results: We achieved the genomewide profiling of copy number aberrations and point mutations in the plasma of the cancer patients. By detecting and quantifying the genomewide aggregated allelic loss and point mutations, we determined the fractional concentrations of tumor-derived DNA in plasma and correlated these values with tumor size and surgical treatment. We also demonstrated the potential utility of this approach for the analysis of complex oncologic scenarios by studying the patient with 2 synchronous cancers. Through the use of multiregional sequencing of tumoral tissues and shotgun sequencing of plasma DNA, we have shown that plasma DNA sequencing is a valuable approach for studying tumoral heterogeneity.

Conclusions: Shotgun DNA sequencing of plasma is a potentially powerful tool for cancer detection, monitoring, and research.

© 2012 American Association for Clinical Chemistry

Comment in

[Plasma-derived tumor DNA analysis at whole-genome resolution.](#)

Swanton C.

[Clin Chem](#). 2013 Jan;59(1):6-8. doi: 10.1373/clinchem.2012.197053. Epub 2012 Oct 31.

PMID: 23115052 No abstract available.

Related information

[MedGen](#)

LinkOut - more resources

Full Text Sources

[Ovid Technologies, Inc.](#)

[Silverchair Information Systems](#)

Other Literature Sources

[The Lens - Patent Citations](#)

[scite Smart Citations](#)

PMID- 23065472

OWN - NLM

STAT- MEDLINE

DCOM- 20130429

LR - 20220321

IS - 1530-8561 (Electronic)

IS - 0009-9147 (Linking)

VI - 59

IP - 1

DP - 2013 Jan

TI - Cancer genome scanning in plasma: detection of tumor-associated copy aberrations, single-nucleotide variants, and tumoral heterogeneity by parallel sequencing.

PG - 211-24

LID - 10.1373/clinchem.2012.196014 [doi]

AB - BACKGROUND: Tumor-derived DNA can be found in the plasma of cancer patients. In this study, we explored the use of shotgun massively parallel sequencing of plasma DNA from cancer patients to scan a cancer genome noninvasively. Four hepatocellular carcinoma patients and a patient with synchronous ovarian cancers were recruited. DNA was extracted from the tumor tissue, and preoperative and postoperative plasma samples of these patients were sequenced with shotgun MPS. RESULTS: We achieved the genomewide profiling of copy number aberrations and point mutations in the plasma of the cancer patients. By detecting and quantifying the genomewide aggregated allelic loss and mutations, we determined the fractional concentrations of tumor-derived DNA in plasma and correlated these values with tumor size and surgical treatment. We also demonstrated the potential utility of this approach for the analysis of complex oncologic scenarios by studying the patient with 2 synchronous cancers. Through the use of multiregional sequencing of tumoral tissues and shotgun sequencing of plasma DNA, we have shown that plasma DNA sequencing is a novel approach for studying tumoral heterogeneity. CONCLUSIONS: Shotgun sequencing of plasma DNA is a potentially powerful tool for cancer detection, monitoring, and research.

CI - © 2012 American Association for Clinical Chemistry

FAU - Chan, K C Allen

AU - Chan KC

AD - Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, New Territories, Hong Kong SAR, China

FAU - Jiang, Peiyong

AU - Jiang P

FAU - Zheng, Yama W L

AU - Zheng YW

FAU - Liao, Gary J W

AU - Liao GJ

FAU - Sun, Hao

AU - Sun H

FAU - Wong, John

AU - Wong J

FAU - Siu, Shing Shun N
AU - Siu SS
FAU - Chan, Wing C
AU - Chan WC
FAU - Chan, Stephen L
AU - Chan SL
FAU - Chan, Anthony T C
AU - Chan AT
FAU - Lai, Paul B S
AU - Lai PB
FAU - Chiu, Rossa W K
AU - Chiu RW
FAU - Lo, Y M D
AU - Lo YM
LA - eng
PT - Journal Article
PT - Research Support, Non-U.S. Gov't
DEP - 20121011
PL - England
TA - Clin Chem
JT - Clinical chemistry
JID - 9421549
SB - IM
CIN - Clin Chem. 2013 Jan;59(1):6-8. PMID: 23115052
MH - Breast Neoplasms/blood/genetics
MH - *DNA Copy Number Variations
MH - Female
MH - *Genes, Neoplasm
MH - *Genome
MH - Humans
MH - Liver Neoplasms/blood/genetics
MH - Ovarian Neoplasms/blood/genetics
MH - Sequence Analysis, DNA/*methods
EDAT- 2012/10/16 06:00
MHDA- 2013/04/30 06:00
CRDT- 2012/10/16 06:00
PHST- 2012/10/16 06:00 [entrez]
PHST- 2012/10/16 06:00 [pubmed]
PHST- 2013/04/30 06:00 [medline]
AID - clinchem.2012.196014 [pii]
AID - 10.1373/clinchem.2012.196014 [doi]
PST - ppublish
S0 - Clin Chem. 2013 Jan;59(1):211-24. doi: 10.1373/clinchem.2012.196014.
Oct 11.

Plasma-Derived Tumor DNA Analysis at Whole-Genome Resolution

Charles Swanton^{1,2*}

The last 2 years have witnessed an extraordinary change in our understanding of the somatic mutational landscape of solid tumors, driven by the impressive large-scale sequencing efforts from such groups as the Cancer Genome Atlas and the Cancer Genome Project. As is so often the case with such ambitious scientific endeavors, the complexities these efforts have revealed raise even greater questions for the future. These studies have exposed the scale of diversity in the mutational landscape among tumors of the same histopathologic subtype. For example, between 1 and 3 genes in high-grade serous ovarian cancer or triple-negative breast cancer are subject to somatic mutations in >10% of patients with the same histopathologic subtype of disease (1, 2). The remainder of the somatic mutational landscape is dominated by genes subject to mutations at much lower frequencies. Low-frequency but potentially targetable events challenge conventional approaches to personalized medicine, both in terms of the health-economic costs in developing therapeutics for such small numbers of patients and in terms of tumor screening to identify such low-frequency somatic events.

Adding to this layer of complexity is the increasing evidence for profound heterogeneity in DNA copy number states, somatic mutational and ploidy profiles that may be distinguished within a single tumor sample (3–5) or between primary and metastatic sites (6–9). Indeed, such subclonal diversity is thought to contribute to adaptation for tumor growth at distant sites of disease (10). Increasing evidence suggests that the emergence of drug resistance in solid tumors may be predetermined by the presence of low-frequency heterogeneous tumor subclones harboring somatic mutations that confer resistance to the targeting anticancer agent (11). Finally, also emerging is clear evidence that the subclonal dynamics of both hematologic and solid tumors change over time [reviewed in (10)] to such an

extent that the subclone that ultimately determines disease outcome may barely be detectable at diagnosis (12).

Such findings have begun to raise profound questions regarding how mutational and DNA copy number profiles can be defined rapidly in individual patients and how future personalized cancer therapeutics can be designed to limit disease progression, predict the early emergence of drug resistance, and resolve the changing nature of a tumor's subclonal dynamics over time (13). Given this emerging evidence for intratumor heterogeneity in somatic mutational status between sites of disease and for the risks associated with biopsying multiple sites of disease (6, 14), it is critical to determine how less-invasive techniques can be established to profile a tumor's somatic mutational load more rapidly and to track its evolution over time. Arguably, the future of oncologic care will rely on more accurate and less invasive diagnostic and predictive tools. Progress in our understanding of drug resistance and tumor metastases will undoubtedly depend on our ability to monitor a tumor's subclonal dynamics and the Darwinian selection of tumor subclones, preferably with just a blood sample rather than multiple sequential biopsies of metastatic sites.

The analysis of circulating free tumor DNA (ctDNA)³ has been primed for some time to overcome the hurdle of invasive biopsies by identifying predetermined somatic mutations via a “liquid biopsy” (15). Until now, however, ctDNA analysis has not been able to resolve distinct copy number events in tumors or to resolve single-nucleotide variants (SNVs) on a genomewide scale. In this cancer-themed issue of *Clinical Chemistry*, Lo and colleagues provide long-awaited evidence that massively parallel sequencing (MPS) analysis of ctDNA can resolve DNA structural aberrations across the entire genome at 1-Mb resolution (16). These authors demonstrate, with 4 patients with hepatocellular carcinoma, the striking representation of copy number events detected in a tumor in events detected in ctDNA, and the almost complete loss of these events from the plasma after tumor resection. In con-

¹ Translational Cancer Therapeutics Laboratory, Cancer Research UK London Research Institute, London, UK; ² UCL Cancer Institute, London, UK.

* Address correspondence to the author at: Translational Cancer Therapeutics Laboratory, Cancer Research UK London Research Institute, 44 Lincoln's Inn Fields, London WC2A 3LY, UK. Fax +44-20-7269-3094; e-mail Charles.swanton@cancer.org.uk.

Received October 2, 2012; accepted October 4, 2012.

Previously published online at DOI: 10.1373/clinchem.2012.197053

³ Nonstandard abbreviations: ctDNA, circulating free tumor DNA; SNV, single-nucleotide variant; MPS, massively parallel sequencing.

trast, only 1% of the sequencing bins for ctDNA isolated from 4 carriers of the hepatitis B virus with no detectable tumor revealed abnormal copy number aberrations, a finding that emphasizes the potential for ctDNA assessment as a screening tool for this disease. Intriguingly, the authors detected ctDNA aberrations that were not present in the single biopsy obtained from an individual's hepatocellular carcinoma. Although Lo and colleagues reflect that such ctDNA aberrations represent reduced specificity for the use of ctDNA analysis for hepatocellular carcinoma, an alternative explanation is that a single biopsy of a tumor may underrepresent the copy number landscape, as has been shown for renal cancer (6). It would therefore be important to assess an equally plausible explanation—that ctDNA analysis may in fact be superior to a single tumor biopsy for resolving the majority of copy number aberrations present in a single tumor.

Remarkably, Lo and colleagues demonstrated through their study of a patient with synchronous breast and ovarian primaries that the structural DNA copy number aberrations detectable in plasma ctDNA were a hybrid of the ovarian and breast primaries. They elegantly demonstrate that the attenuation of DNA copy number signals in ctDNA after sequential, same-day surgical resection of the breast primary and bilateral ovarian cancers directly mirrors the single-nucleotide polymorphism/comparative genomic hybridization copy number profiles derived from the surgically resected tumors. Although the breast primary was only 3 cm in size and contributed an estimated 2.1% of the total ctDNA, the authors showed that this analytically sensitive technique was still able to detect breast cancer-specific copy number aberrations—such as a deletion in chromosome 6p and amplifications on chromosomes 1q, 7p, and 15q—that were present only in the breast primary and disappeared soon after breast surgery.

These data provide conclusive evidence that the use of plasma-derived ctDNA for unbiased analysis of the structural genomic landscape of tumors over time is now fit for implementation in the research setting. Given the clinical risks of repeat tumor biopsies and the challenges associated with inter- and intratumor heterogeneity, this landmark study will enable rapid progress in our understanding of tumor evolution over time.

Perhaps most importantly, the authors have used ctDNA analysis to explore the phenomenon of intratumoral heterogeneity in ovarian cancer. Although 4 regions of the bilateral ovarian tumors had almost identical copy number profiles, the authors found SNVs specific to a single region, shared SNVs within a single ovarian cancer, and ubiquitous SNVs present bilaterally. Using a mass spectrometry-based iPLEX analyti-

cal method, the authors conclusively identified 95% of the 67 SNVs selected for validation in the tumor DNA. The authors then leveraged this analysis to define the presence of these SNVs in the MPS data from ctDNA before and after surgery. In their analysis of the ubiquitous SNVs, the authors estimated that these SNVs accounted for 46% of the ctDNA before surgery and 0.18% after bilateral oophorectomy. When the authors focused their attention on the SNVs shared by one or the other of the ovarian tumors, they found that the degree of a tumor's SNV representation in the ctDNA mirrored the size of the ovarian tumor mass on the side from which these SNVs derived, suggesting that ctDNA may be a useful surrogate of residual disease.

These data have important implications. First, it is apparent that tumor heterogeneity poses a substantial problem to the use of ctDNA focused on tumor-specific mutations as a tool to predict early relapse or to monitor the emergence of drug resistance, if the focus is on somatic events that may not be ubiquitous throughout the tumor mass. Second, these data support the use of unbiased shotgun MPS approaches to ctDNA analysis—particularly in the context of studying the subclonal dynamics of the disease through therapy—for attempting to resolve the emergence of distinct tumor subclones harboring low-frequency somatic events conferring drug resistance. With the future acceleration of DNA sequencing and data analysis, the shotgun MPS approach to ctDNA analysis has the potential to provide a powerful tool for deciphering the Darwinian mechanisms of tumor evolution and to keep pace with tumor subclonal dynamics through the disease course.

Author Contributions: All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.

Authors' Disclosures or Potential Conflicts of Interest: No authors declared any potential conflicts of interest.

References

1. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* 2011;474:609–15.
2. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* 2012;486:395–9.
3. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, et al. The life history of 21 breast cancers. *Cell* 2012;149:994–1007.
4. Navin N, Krasnitz A, Rodgers L, Cook K, Meth J, Kendall J, et al. Inferring tumor progression from genomic heterogeneity. *Genome Res* 2010;20:68–80.
5. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, et al. Tumour evolution inferred by single-cell sequencing. *Nature* 2011;472:90–4.

6. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* 2012;366:883–92.
7. Wu X, Northcott PA, Dubuc A, Dupuy AJ, Shih DJ, Witt H, et al. Clonal selection drives genetic divergence of metastatic medulloblastoma. *Nature* 2012;482:529–33.
8. Campbell PJ, Yachida S, Mudie LJ, Stephens PJ, Pleasance ED, Stebbings LA, et al. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* 2010;467:1109–13.
9. Shah SP, Morin RD, Khattra J, Prentice L, Pugh T, Burleigh A. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* 2009;461:809–13.
10. Swanton, C. Intratumor heterogeneity: evolution through space and time. *Cancer Res* 2012;72:4875–82.
11. Su KY, Chen HY, Li KC, Kuo ML, Yang JC, Chan WK, et al. Pretreatment epidermal growth factor receptor (EGFR) T790M mutation predicts shorter EGFR tyrosine kinase inhibitor response duration in patients with non-small-cell lung cancer. *J Clin Oncol* 2012;30:433–40.
12. Keats JJ, Chesi M, Egan JB, Garbitt VM, Palmer SE, Braggio E, et al. Clonal competition with alternating dominance in multiple myeloma. *Blood* 2012;120:1067–76.
13. Yap TA, Gerlinger M, Futreal PA, Pusztai L, Swanton C. Intratumour heterogeneity: seeing the wood for the trees. *Sci Transl Med* 2012;4:127.
14. Chen ZY, Zhong WZ, Zhang XC, Su J, Yang XN, Chen ZH, et al. EGFR mutation heterogeneity and the mixed response to EGFR tyrosine kinase inhibitors of lung adenocarcinomas. *Oncologist* 2012;17:978–85.
15. Forsheo T, Murtaza M, Parkinson C, Gale D, Tsui DW, Kaper F, et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci Transl Med* 2012;4:136–68.
16. Chan KCA, Jiang P, Zheng YWL, Liao GJW, Sun H, Wong J, Siu SSN, et al. Cancer genome scanning in plasma: detection of tumor-associated copy number aberrations, single-nucleotide variants, and tumoral heterogeneity by massively parallel sequencing. *Clin Chem* 2013;59:211–24.

FULL TEXT LINKS



Editorial > [Clin Chem](#). 2013 Jan;59(1):6-8. doi: 10.1373/clinchem.2012.197053.

Epub 2012 Oct 31.

Plasma-derived tumor DNA analysis at whole-genome resolution

[Charles Swanton](#)

PMID: 23115052 DOI: [10.1373/clinchem.2012.197053](#)

No abstract available

Comment on

Cancer genome scanning in plasma: detection of tumor-associated copy number aberrations, single-nucleotide variants, and tumoral heterogeneity by massively parallel sequencing.

Chan KC, Jiang P, Zheng YW, Liao GJ, Sun H, Wong J, Siu SS, Chan WC, Chan SL, Chan AT, Lai PB, Chiu RW, Lo YM.

Clin Chem. 2013 Jan;59(1):211-24. doi: 10.1373/clinchem.2012.196014. Epub 2012 Oct 11.

PMID: 23065472

LinkOut - more resources

Full Text Sources

[Ovid Technologies, Inc.](#)

[Silverchair Information Systems](#)

Other Literature Sources

[scite Smart Citations](#)

PMID- 23115052
OWN - NLM
STAT- MEDLINE
DCOM- 20130429
LR - 20190722
IS - 1530-8561 (Electronic)
IS - 0009-9147 (Linking)
VI - 59
IP - 1
DP - 2013 Jan
TI - Plasma-derived tumor DNA analysis at whole-genome resolution.
PG - 6-8
LID - 10.1373/clinchem.2012.197053 [doi]
FAU - Swanton, Charles
AU - Swanton C
LA - eng
PT - Comment
PT - Editorial
DEP - 20121031
PL - England
TA - Clin Chem
JT - Clinical chemistry
JID - 9421549
SB - IM
CON - Clin Chem. 2013 Jan;59(1):211-24. PMID: 23065472
MH - *DNA Copy Number Variations
MH - Female
MH - *Genes, Neoplasms
MH - *Genome
MH - Humans
MH - Sequence Analysis, DNA/*methods
EDAT- 2012/11/02 06:00
MHDA- 2013/04/30 06:00
CRDT- 2012/11/02 06:00
PHST- 2012/11/02 06:00 [entrez]
PHST- 2012/11/02 06:00 [pubmed]
PHST- 2013/04/30 06:00 [medline]
AID - clinchem.2012.197053 [pii]
AID - 10.1373/clinchem.2012.197053 [doi]
PST - ppublish
S0 - Clin Chem. 2013 Jan;59(1):6-8. doi: 10.1373/clinchem.2012.197053. Epub
31.

Try the [Copyright Public Records System \(CPRS\)](#) pilot with enhanced search features and filters.

[Help](#)[Search](#)[History](#)[Titles](#)[Start Over](#)

Public Catalog

Copyright Catalog (1978 to present)

Search Request: Left Anchored Title = clinical chemistry

Search Results: Displaying 92 of 424 entries

[previous](#)[next](#)[Labeled View](#)

Clinical Chemistry.

Type of Work: Serial

Registration Number / Date: TX0007648861 / 2013-01-31

Application Title: Clinical Chemistry

Title: Clinical Chemistry. [Published: 2012-12-31. Issue: vol. 59, no. 1, January 2013]

Serial Publication Year: 2012

ISSN: 0009-9147

Frequency: Monthly.

Copyright Claimant: American Association for Clinical Chemistry, Inc. Address: 1850 K Street, Suite 625, Washington, DC, 20006, United States.

Nation of First Publication: United States

Authorship on Application: American Association for Clinical Chemistry, employer for hire; Domicile: United States; Citizenship: United States. Authorship: contribution(s) to a collective work.

Issues Registered: vol. 59, no. 1, January 2013. Created 2012; Pub. 2012-12-31. Reg. 2013-01-31; TX0007648861

Names: [American Association for Clinical Chemistry](#)
[American Association for Clinical Chemistry, Inc](#)

[previous](#)[next](#)

Save, Print and Email ([Help Page](#))

Select Download Format

Enter your email address:

[Help](#) [Search](#) [History](#) [Titles](#) [Start Over](#)

[Contact Us](#) | [Request Copies](#) | [Get a Search Estimate](#) | [Frequently Asked Questions \(FAQs\) about Copyright](#) | [Copyright Office Home Page](#) | [Library of Congress Home Page](#)

Development of Personalized Tumor Biomarkers Using Massively Parallel Sequencing

Rebecca J. Leary,¹ Isaac Kinde,¹ Frank Diehl,¹ Kerstin Schmidt,¹ Chris Clouser,² Cisilya Duncan,² Alena Antipova,² Clarence Lee,² Kevin McKernan,² Francisco M. De La Vega,³ Kenneth W. Kinzler,¹ Bert Vogelstein,¹ Luis A. Diaz Jr.,¹ Victor E. Velculescu^{1*}

(Published 24 February 2010; Volume 2 Issue 20 20ra14)

Clinical management of human cancer is dependent on the accurate monitoring of residual and recurrent tumors. The evaluation of patient-specific translocations in leukemias and lymphomas has revolutionized diagnostics for these diseases. We have developed a method, called personalized analysis of rearranged ends (PARE), which can identify translocations in solid tumors. Analysis of four colorectal and two breast cancers with massively parallel sequencing revealed an average of nine rearranged sequences (range, 4 to 15) per tumor. Polymerase chain reaction with primers spanning the breakpoints was able to detect mutant DNA molecules present at levels lower than 0.001% and readily identified mutated circulating DNA in patient plasma samples. This approach provides an exquisitely sensitive and broadly applicable approach for the development of personalized biomarkers to enhance the clinical management of cancer patients.

INTRODUCTION

A nearly universal feature of human cancer is the widespread rearrangement of chromosomes as a result of chromosomal instability (1). Such structural alterations begin to occur at the earliest stages of tumorigenesis and persist throughout tumor development. The consequences of chromosomal instability can include copy number alterations (duplications, amplifications, and deletions), inversions, insertions, and translocations (2). Historically, the ability to detect such alterations has been limited by the resolution of genetic analyses. However, a number of more recent approaches, including high-density oligonucleotide arrays and high-throughput sequencing, have allowed detection of changes at much higher resolution (3–15).

Tumor-specific (somatic) chromosomal rearrangements have the potential to serve as highly sensitive biomarkers for tumor detection. Such alterations are not present in normal cells and should be exquisitely specific. Rearrangement-associated biomarkers therefore offer a reliable measure that would be useful for monitoring tumor response to specific therapies, detecting residual disease after surgery, and long-term clinical management. Recurrent somatic structural alterations, such as those involving the BCR-ABL oncogene (the target of the Philadelphia chromosome translocation), immunoglobulin genes, T cell receptor genes, and the retinoic acid receptor α gene, have been shown to be useful as diagnostic markers in certain hematopoietic malignancies (16–20). However, recurrent structural alterations do not generally occur in most solid tumors. We reasoned that any structural alteration identified in an individual's tumor could be used for analogous purposes, whether it was found in tumors of the same type in other individuals and whether it was a “driver”—causing a selective growth advantage—or a “passenger.” We describe herein our efforts to implement this concept in representative examples of common solid tumors.

RESULTS

Description of the approach

The PARE (personalized analysis of rearranged ends) approach, shown schematically in Fig. 1, requires the identification of patient-specific rearrangements in tumor samples. To determine the feasibility of identifying such alterations using next-generation sequencing approaches, we initially analyzed four tumor samples (two colon and two breast tumors) and their matched normal tissue samples using the Applied Biosystems SOLiD System (Table 1). Genomic DNA from each sample was purified, sheared, and used to generate libraries with mate-paired tags ~1.4 kb apart. Libraries were digitally amplified by emulsion polymerase chain reaction (PCR) on magnetic beads (21), and 25-bp mate-paired tags were sequenced using the sequencing-by-ligation approach (15, 22). An average of 198.1 million 25-bp reads was obtained for each sample, where each read aligned perfectly and was uniquely localized in the reference human genome (hg18), resulting in 4.95 Gb of mappable sequence per sample. An average of 40 million mate-paired reads, where both tags were perfectly mapped to the reference human genome, was obtained for each sample. The total amount of genome base pairs covered by the mate-paired analysis (that is, distance between mate-paired tags $\times$ number of mate-paired tags) was 53.6 Gb per sample, or an 18-fold physical coverage of the human genome.

Identification of somatic rearrangements

Two methods were used to identify somatic rearrangements from these data (fig. S1). The first approach involved searching for tags whose mate pairs were derived from different chromosomes (inter-chromosomal rearrangements). The high physical coverage of breakpoints provided by the ~40 million mate-paired sequences per sample (Table 1) suggested that a large fraction of such translocations could be identified. End sequences from such mate-paired tags were grouped into 1-kb bins, and those bin pairs that were observed at least five times were analyzed further. The requirement for five or more occurrences minimized the chance that the presumptive fusion sequences

¹Ludwig Center for Cancer Genetics and Therapeutics and Howard Hughes Medical Institute, Johns Hopkins Kimmel Cancer Center, Baltimore, MD 21231, USA. ²Life Technologies, Beverly, MA 01915, USA. ³Life Technologies, Foster City, CA 94404, USA. *To whom correspondence should be addressed. E-mail: velculescu@jhmi.edu

represent incorrect mapping to the reference genome or artifacts of library construction. Comparison with SOLiD libraries made from the matched normal samples reduced the possibility that the fusion sequences represented rare germline variants rather than somatic events.

Fig. 1. Schematic of PARE approach. The method is based on next-generation mate-paired analysis of resected tumor DNA to identify individualized tumor-specific rearrangements. Such alterations are used to develop PCR-based quantitative analyses for personalized tumor monitoring of plasma samples or other bodily fluids.

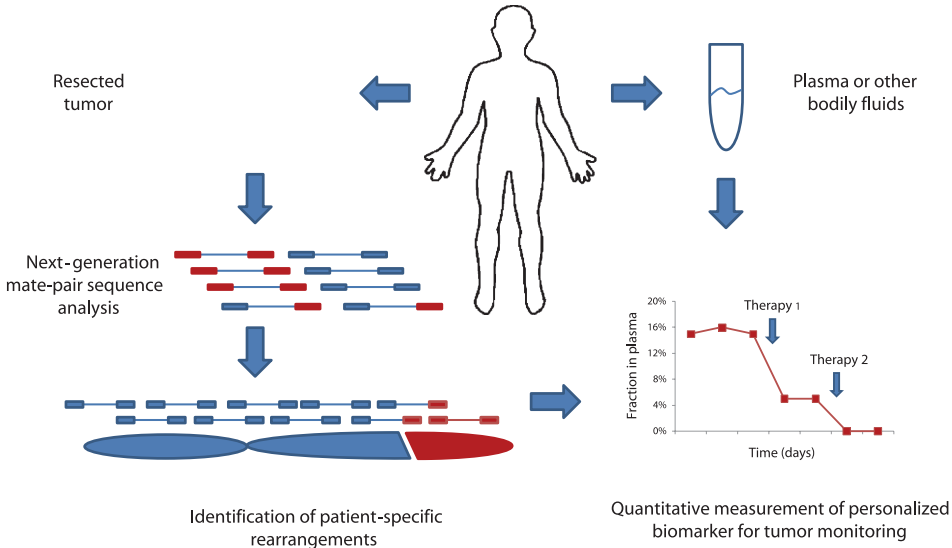


Table 1. Summary of mate-paired tag libraries.

Samples	Single tag analyses				Mate-paired tag analyses			
	Number of beads*	Number of tags matching human genome	Total bases sequenced (bp)	Expected coverage per 3-kb bin	Number of mate-paired tags matching human genome	Distance between mate-paired tags (bp)	Total physical coverage by mate-paired tags (bp)	Expected genome coverage
Colon cancer								
Co108 tumor	526,209,780	121,527,707	3,038,192,675	122	21,899,809	1371	30,024,693,714	10.0
Co108 normal	328,599,033	86,032,253	2,150,806,325	86	11,694,361	1254	14,665,530,804	4.9
Co84 tumor	677,137,128	256,065,437	6,401,635,925	256	58,678,410	1488	87,292,060,006	29.1
Co84 normal	486,663,520	218,280,146	5,457,003,650	218	59,019,031	1384	81,690,396,379	27.2
Hx402 tumor	523,745,015	198,342,749	4,958,568,725	198	43,457,431	1629	70,789,547,653	23.6
Hx403 tumor	475,658,760	164,061,938	4,101,548,450	164	37,123,395	1705	63,295,388,475	21.1
Breast cancer								
B7 tumor	840,979,999	281,027,274	7,025,681,850	281	27,548,989	1220	33,604,662,404	11.2
B7 normal	705,704,265	253,482,262	6,337,056,550	253	57,878,644	1404	81,271,654,770	27.1
B5 tumor	444,249,217	147,612,941	3,690,323,525	148	29,961,045	1193	35,730,144,651	11.9
B5 normal	549,237,156	220,669,795	5,516,744,875	221	53,611,974	1205	64,591,276,025	21.5

*Number of beads corresponds to the number of magnetic beads containing clonally amplified DNA fragments and represents the maximal number of raw sequence reads for each run.

Table 2. Summary of rearrangements identified in tumor samples.

Sample	Rearrangement type		Total rearrangements	Tested rearrangements	Confirmed somatic rearrangements
	Intrachromosomal	Interchromosomal			
Tumor and normal libraries					
B5	7	4	11	7	5 (71%)
B7	17	4	21	16	15 (94%)
Co84	0	7	7	6	4 (67%)
Co108	6	12	18	13	11 (85%)
Tumor libraries					
Hx402	7	2	9	9	4 (44%)
Hx403	17	0	17	12	7 (58%)

would be expected to contain novel junctions not present in the human genome. To identify somatic copy number gains, losses, high-amplitude amplifications, and homozygous deletions, we grouped tags into nonoverlapping 3-kb bins. Normalized tag densities, defined as the number of tags per bin divided by average number of tags per bin, were determined for all 3-kb bins in each sample. Bins that displayed tag density ratios of >1.75 or <0.25 in two or more normal tissue samples (corresponding to $<6\%$ of all bins) were discarded from the analysis. This eliminated confounding regions of common germline copy number variation and resulted in 892,567 bins that were analyzed in each tumor sample. Comparison of 256 million reads from colorectal tumor sample Co84 with Illumina arrays containing ~1 million single-nucleotide polymorphism (SNP) probes and with a ~1 million Digital Karyotyping tag library obtained with Sanger sequencing showed high concordance for copy number alterations among the three platforms (fig. S2 and table S1). With the higher resolution afforded by the SOLiD data, we were able to identify additional copy number changes not detected with the other methods (table S2). Boundary regions of copy number alteration were analyzed to identify mate-paired tags corresponding to rearranged DNA sequences. These included fusion of DNA sequences that have inappropriate spacing, order, or orientation on the same chromosome (intrachromosomal rearrangements), or inappropriate joining of sequences from different chromosomes (interchromosomal rearrangements).

Through these two approaches, we identified 57 regions containing putative somatic rearrangements, with an average of 14 rearrangements per sample (Table 2). Of these, an average of seven represented interchromosomal rearrangements and seven represented intrachromosomal rearrangements. For confirmation, we designed primers to 42 of the paired-end regions and used them for PCR spanning the putative breakpoints. Thirty-five of these (83%) yielded PCR products

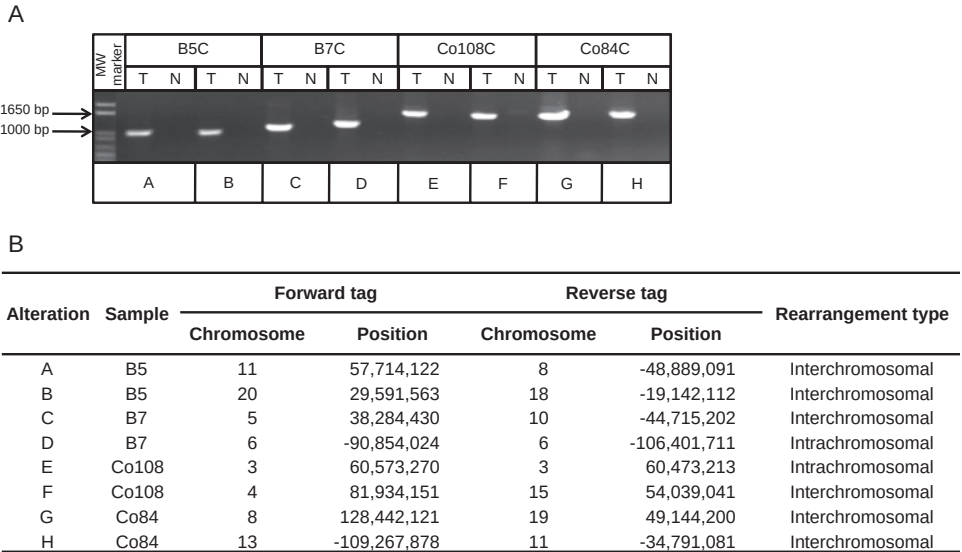


Fig. 2. Detection of tumor-specific rearrangements in breast and colorectal cancers. Two representative rearrangements are shown for each tumor sample. **(A)** PCR amplification across breakpoint regions. MW, molecular weight; T, tumor; N, normal. **(B)** Genomic coordinates for a representative mate pair of each rearrangement.

of the expected size in the tumor samples but not in the normal samples (Fig. 2 and table S3). Sanger sequencing of seven PCR products confirmed the rearrangements in all cases tested. Although there was variation in the number of detected alterations per sample (range, 7 to 21), all four tumor samples were found to have at least four bona fide somatic rearrangements through this approach.

Further examination revealed that rearrangements could be readily identified with high confidence even in the absence of data from matched normal DNA by using the copy number and mate-pair coupled approach. Elimination of analysis of the matched normal would reduce the cost and simplify the identification of rearrangements. To test this strategy, we analyzed two additional tumor samples (Hx402 and Hx403) through the SOLiD approach but without generation of matching normal DNA libraries. We found that it was possible to identify putative rearrangements resulting in interchromosomal and intrachromosomal rearrangements at the border of copy number variations with high specificity even in the absence of a matched

Downloaded from https://www.science.org on October 17, 2022

normal library. We were able to identify 11 confirmed somatic alterations (4 in Hx402 and 7 in Hx403) out of 21 candidate changes tested (table S3).

Development of PARE biomarkers from rearranged sequences

Each of the rearranged sequences identified through PARE was unique, as no identical rearrangement was found in any of the other five tumor samples. To determine the utility of these rearranged sequences to serve as potential biomarkers, we designed PCR assays to detect them in the presence of increasing amounts of normal DNA. These conditions simulate detection of tumor DNA from patient blood or other bodily fluids, where tumor DNA constitutes a minority of total DNA. PCR products representing a rearranged region from each of the six dilutions of tumor DNA could be identified, even in mixtures of DNA containing 1 cancer genome equivalent among ~390,000 normal genome equivalents (Fig. 3). Furthermore, no background PCR products were discernible when DNA from normal tissues was used as control.

To determine whether the rearranged sequences could actually be detected in clinical samples, we evaluated circulating DNA from plasma samples of patients Hx402 and Hx403. The sample from patient Hx403 was obtained before surgery, whereas the samples from patient Hx402 were obtained before and after surgery. A chromosome 4:8 translocation associated with an amplification was used in tumor Hx402, and an intrachromosomal rearrangement associated with a homozygous deletion of chromosome 16 was used in tumor Hx403. PCR amplification of plasma DNA using primers spanning the breakpoints produced products of the expected sizes only in the plasma samples from patients with disease and not in plasma from healthy controls (Fig. 4A). Sequencing of the PCR products from plasma DNA identified the identical breakpoints observed in the tumor DNA samples.

We next determined whether circulating tumor-derived rearranged DNA could be quantitatively measured in patient plasma. Digital PCR analyses of plasma samples from patient Hx402 indicated that rearranged tumor DNA was detectable at all six time points analyzed throughout the course of treatment (Fig. 4B). The fraction of mutant

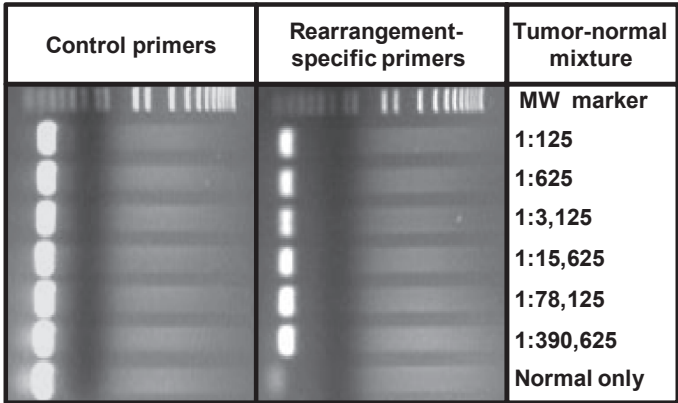


Fig. 3. Detection of tumor-specific rearrangements in mixtures of tumor and normal DNA. Decreasing amounts of tumor DNA were mixed with increasing amounts of normal tissue DNA (300 ng total) and were used as template molecules for PCR using chromosome 4:8 translocation-specific primers or chromosome 3 control primers (see Materials and Methods for additional information).

DNA present in plasma samples was 37% just before surgery but decreased to 14% 1 day after resection of the primary tumor. The mutant DNA fraction decreased further after chemotherapy and subsequent removal of metastatic lesions from the right lobe of the liver. However, the fraction of mutant tumor DNA did not reach zero (remaining at 0.3% at day 137), consistent with the fact that this patient had residual metastatic lesions in the remaining left lobe of the liver.

DISCUSSION

These results demonstrate that massively parallel sequencing can be used to develop personalized biomarkers based on somatic rearrangements. We were able to identify tumor-specific markers in each of the six breast and colorectal cancer cases analyzed. Moreover, we demon-

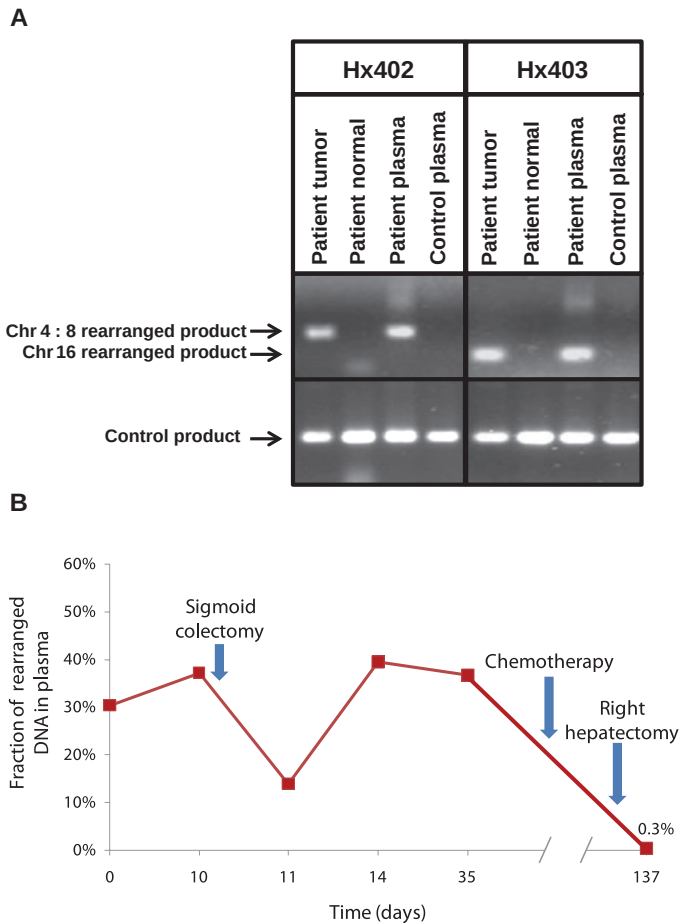


Fig. 4. Detection of tumor-specific rearrangements in plasma of cancer patients. **(A)** The identified chromosome 4:8 and 16 rearrangements were used to design PCR primers spanning breakpoints and to amplify rearranged DNA from tumor tissue and plasma from patients Hx402 and Hx403, respectively. A plasma sample from an unrelated healthy individual was used as a control for both rearrangements. **(B)** Plasma samples from patient Hx402 were analyzed at different time points using digital PCR to determine the fraction of genomic equivalents of plasma DNA containing the chromosome 4:8 rearrangement. The fraction of rearranged DNA at day 137 was 0.3%, consistent with residual metastatic lesions present in the remaining lobe of the liver.

Downloaded from https://www.science.org on October 17, 2022

strated that the identified breakpoints can be used to detect tumor DNA in the presence of large quantities of normal DNA and in patient plasma. These results highlight the sensitivity and specificity of the approach and suggest broad clinical utility of the PARE method.

Virtually all tumors of clinical consequence are thought to have rearranged DNA sequences resulting from translocations and copy number alterations, and these sequences are not present in normal human plasma or nontumor tissues. A recent genome-wide analysis of 24 breast cancers showed that all analyzed samples contained at least one genomic rearrangement that could be detected by next-generation sequencing (24). From a technical perspective, PARE-derived clinical assays should have no false positives: The PCR amplification of aberrant fusions of DNA sequences that are normally thousands of base pairs apart or on different chromosomes should not occur in nontumor DNA. In contrast, approaches that rely on monitoring of residual disease by analysis of somatic single-base alterations in specific genes are limited by polymerase error rates at the bases of interest (25). The PCR process generates background single-base mutations that are identical to bona fide mutations but does not generate false-positive rearrangements with carefully chosen primers. Because of the higher signal-to-noise ratio thereby obtained, PARE theoretically permits more sensitive monitoring of tumor burden.

The PARE approach, however, is not without limitations. Although somatic alterations in oncogenes and tumor suppressor genes persist throughout the clonal evolution of a tumor, it is conceivable that some rearranged sequences could be lost during tumor progression. The identification of several PARE biomarkers, each specific for different chromosomal regions, would mitigate this concern, as it is unlikely that all such markers would be lost in any particular patient. Another limitation is the cost of identifying a patient-specific alteration. In this prototype study, we obtained an average of 194.7 million reads per patient, resulting in ~200 tags in each 3-kb bin. The current cost for such an assay is ~\$5000, which is expensive for general clinical use. This cost is a consequence of the high physical coverage and the inefficiencies associated with stringent mapping of 25-bp sequence data to the human genome. As read quality and length continue to improve, less stringent mapping criteria and lower physical coverage will permit analyses similar to those in this study but with substantially less sequencing effort. Moreover, the cost of massively parallel sequencing, which has decreased substantially over the last 2 years, continues to spiral downwards. Finally, there are clinical settings where the fraction of any DNA from tumors, including rearranged sequences, in the patient plasma is exceedingly small and undetectable. To be detectable by PARE, there must be at least one rearrangement template molecule in the plasma sample analyzed. When disease burden is this light, PARE may yield false-negative results. Larger studies will be needed to determine the clinical utility of PARE and its prognostic capabilities.

Despite these caveats, there are numerous potential applications of PARE. These include the more accurate identification of surgical margins free of tumor and the analysis of regional lymph nodes as well as the measurement of circulating tumor DNA after surgery, radiation, or chemotherapy. Short-term monitoring of circulating tumor DNA may be particularly useful in the testing of new drugs, as it could provide an earlier indication of efficacy than is possible through conventional diagnostic methods such as computed tomography (CT) scanning. Given current enthusiasm for the personalized manage-

ment of cancer patients, PARE affords a timely method for uniquely sensitive and specific tumor monitoring.

MATERIALS AND METHODS

Clinical samples and cell lines

DNA samples were obtained from early-passage xenografts and cell lines of breast and colorectal cancers as described (26). Normal DNA samples were obtained from matched normal tissue. Plasma samples were collected from colorectal cancer patients Hx402 and Hx403 and from an unrelated normal control. All samples were obtained in accordance with the Health Insurance Portability and Accountability Act.

Digital Karyotyping and Illumina BeadChip arrays

A Digital Karyotyping library for colorectal cancer cell line Co84C was constructed as previously described (6). In summary, 17-bp genomic DNA tags were generated using the *Nla*III and *Sac*I restriction enzymes. The experimental tags obtained were concatenated, cloned, and sequenced. Previously described software was used to extract the experimental tags from the sequencing data. The sequences of the experimental tags were compared to the predicted virtual tags extracted from the human genome reference sequence. Amplifications were identified using sliding windows of variable sizes, and windows with tag density ratios of ≥ 6 were considered to represent amplified regions.

The Illumina Infinium II Whole Genome Genotyping Assay using the BeadChip platform was used to analyze the colorectal cancer cell line Co84C at 317K SNP loci from the Human HapMap collection. This assay is a two-step procedure: First the sample is hybridized to a 50-nucleotide oligo, and then the SNP position is interrogated by a two-color fluorescent single-base extension. Image files and data normalization were processed as previously described (10). Amplifications were defined as regions having at least 1 SNP with a LogR ratio of ≥ 1.4 , at least 1 in 10 SNPs with a LogR ratio of ≥ 1 , and an average LogR ratio of the entire region of ≥ 0.9 .

SOLiD library preparation and sequencing

Mate-pair libraries were generated for the SOLiD platform as described (15). In brief, genomic DNA was sheared into ~1.4-kb fragments and used as template in emulsion PCR. Fragments were coupled to beads via an adapter sequence and clonally amplified. A 3' modification of the DNA fragments allowed for covalent attachment to a slide. Sequencing primers hybridized to the adapter sequence and four fluorescently labeled di-base probes were used in ligation-based sequencing. Each nucleotide is sequenced twice in two different ligation reactions, resulting in two-base encoding, which has been shown to reduce sequencing artifacts.

Sequence data were mapped to the human genome reference sequence (hg18) using the Corona SOLiD software pipeline. All 25-bp tags (for both individual tag and mate-paired tag analyses) were required to match the reference genome uniquely and without mismatches.

Analysis of single tags for copy number alterations

The SOLiD tags were filtered and the remaining tags were grouped by genomic position in nonoverlapping 3-kb bins. A tag density ratio was calculated for each bin by dividing the number of tags observed in the bin by the average number of tags expected to be in each bin (on the basis of the total number of tags obtained for chromosomes

1 to 22 for each library divided by 849,434 total bins). The tag density ratio thereby allowed a normalized comparison between libraries containing different numbers of total tags. A control group of SOLiD libraries made from the four matched normal samples from Table 1 and two additional normal samples [CEPH (Centre d'Etude du Polymorphisme Humain) samples NA07357 and NA18507] was used to define areas of germline copy number variation or that contained a large fraction of repeated or low-complexity sequences. Any bin where at least two of the normal libraries had a tag density ratio of <0.25 or >1.75 was removed from further analysis.

Homozygous deletions were identified as three or more consecutive bins with tag ratios of <0.25 and at least one bin with a tag ratio of <0.005 . Amplifications were identified as three or more consecutive bins with tag ratios of >2.5 and at least one bin with a tag ratio of >6 . Single-copy gains and losses were identified through visual inspection of tag density data for each sample.

Analysis of mate-paired tags

Mate-paired tags mapping the reference genome uniquely and without mismatches were analyzed for aberrant mate-pair spacing, orientation, and ordering and categorized in 13 three-letter data formats (27). Mate pairs from the same chromosome that map at appropriate distances (~ 1.4 kb) and in the appropriate orientation and ordering are categorized as AAA. Mate pairs mapping to different chromosomes are categorized as C**. For the analysis of translocations of the PARE approach, we focused on C** mate pairs, whereas for analysis of rearrangements adjacent to copy number alterations, we chose all non-AAA (including C**) mate pairs for further analysis.

PARE identification and confirmation of candidate rearrangements

To identify candidate translocations, we grouped C** mate-pair tags in 1-kb bins and looked for bin pairs that were observed five or more times in the tumor sample but that were not observed in the matched normal sample. For identification of candidate rearrangements associated with copy number alterations, we analyzed the 10-kb boundary regions of amplifications, homozygous deletions, or lower copy gains and losses for neighboring non-AAA tags observed more than two times in the tumor but not in the matched normal sample. In the case of Hx402 and Hx403, the analysis of rearrangements adjacent to copy number alterations was performed in the absence of SOLiD libraries from normal tissue.

Mate-pair tag sequences associated with a candidate rearrangement were used as target sequences for primer design using Primer3 (28). When primers could not be designed from tag sequences alone, adjacent genomic sequence up to 100 bp was used for primer design. The observed rearranged tag ordering and orientation was used for Primer3 queries. Primers were used for PCR on tumor and matched normal samples as previously described (26). The candidate rearrangement was confirmed if a PCR product of the expected size was seen in the tumor but not in the matched normal sample. Sanger sequencing of PCR products was used to identify sequence breakpoint in a subset of cases.

Detection of PARE biomarker in human plasma

To determine the sensitivity of rearranged biomarkers in the presence of normal DNA, we used serial dilutions of tumor-normal DNA mixtures as templates for PCR using primers for the chromosome 4:8 trans-

location in Hx402. The tumor DNA dilution began at 1:125 tumor-normal and continued as a one-in-five serial dilution until reaching 1:390,625 tumor-normal mixture. PCR was performed for each of the six tumor-normal DNA mixtures and for the normal DNA control using translocation-specific primers as well as control primers from chromosome 3.

Human plasma samples (1 ml) were obtained from patients Hx402 and Hx403 and from a control individual, and DNA was purified as described (29). Whole-genome amplification of plasma DNA was performed by ligation of adaptor sequences and PCR amplification with universal primers from the Illumina Genomic DNA Sample Prep kit.

Primers designed to amplify <200 -bp fragments spanning each PARE rearrangement were used in PCR from total plasma DNA using patient or control samples. Digital PCR of plasma DNA dilutions from patient Hx402 using rearrangement-specific and control primers was used to quantitate the fraction-mutated DNA molecules.

SUPPLEMENTARY MATERIAL

www.sciencetranslationalmedicine.org/cgi/content/full/2/20/20ra14/DC1

Table S1. Comparison of SOLiD sequencing, Illumina SNP arrays, and Digital Karyotyping for analysis of copy number alterations.

Table S2. Putative copy number alterations identified by SOLiD sequencing in Co84 that were not identified by Illumina SNP arrays or Digital Karyotyping.

Table S3. Confirmed somatic rearrangements in breast and colorectal cancer samples.

Fig. S1. Flow chart of approach used to identify rearranged sequences.

Fig. S2. Comparison of Digital Karyotyping, Illumina SNP array, and SOLiD sequencing results on chromosome 8.

REFERENCES AND NOTES

1. C. Lengauer, K. W. Kinzler, B. Vogelstein, Genetic instabilities in human cancers. *Nature* **396**, 643–649 (1998).
2. F. Mitelman, B. Johansson, F. Mertens, The impact of translocations and gene fusions on cancer causation. *Nat. Rev. Cancer* **7**, 233–245 (2007).
3. D. Pinkel, R. Seagraves, D. Sudar, S. Clark, I. Poole, D. Kowbel, C. Collins, W. L. Kuo, C. Chen, Y. Zhai, S. H. Dairkee, B. M. Ljung, J. W. Gray, D. G. Albertson, High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.* **20**, 207–211 (1998).
4. R. Lucito, J. Healy, J. Alexander, A. Reiner, D. Esposito, M. Chi, L. Rodgers, A. Brady, J. Sebat, J. Troge, J. A. West, S. Rostan, K. C. Nguyen, S. Powers, K. Q. Ye, A. Olshen, E. Venkatraman, L. Norton, M. Wigler, Representational oligonucleotide microarray analysis: A high-resolution method to detect genome copy number variation. *Genome Res.* **13**, 2291–2305 (2003).
5. D. A. Peiffer, J. M. Le, F. J. Steemers, W. Chang, T. Jenniges, F. Garcia, K. Haden, J. Li, C. A. Shaw, J. Belmont, S. W. Cheung, R. M. Shen, D. L. Barker, K. L. Gunderson, High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res.* **16**, 1136–1148 (2006).
6. T. L. Wang, C. Maierhofer, M. R. Speicher, C. Lengauer, B. Vogelstein, K. W. Kinzler, V. E. Velculescu, Digital karyotyping. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16156–16161 (2002).
7. T. L. Wang, L. A. Diaz Jr., K. Romans, A. Bardelli, S. Saha, G. Galizia, M. Choti, R. Donehower, G. Parmigiani, I.-M. Shih, C. Iacobuzio-Donahue, K. W. Kinzler, B. Vogelstein, C. Lengauer, V. E. Velculescu, Digital karyotyping identifies thymidylate synthase amplification as a mechanism of resistance to 5-fluorouracil in metastatic colorectal cancer patients. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3089–3094 (2004).
8. C. Di, S. Liao, D. C. Adamson, T. J. Parrett, D. K. Broderick, Q. Shi, C. Lengauer, J. M. Cummins, V. E. Velculescu, D. W. Fults, R. E. McLendon, D. D. Bigner, H. Yan, Identification of OTX2 as a medulloblastoma oncogene whose product can be targeted by all-trans retinoic acid. *Cancer Res.* **65**, 919–924 (2005).
9. K. Nakayama, N. Nakayama, B. Davidson, H. Katabuchi, R. J. Kurman, V. E. Velculescu, I.-M. Shih, T. L. Wang, Homozygous deletion of MKK4 in ovarian serous carcinoma. *Cancer Biol. Ther.* **5**, 630–634 (2006).
10. R. J. Leary, J. C. Lin, J. Cummins, S. Boca, L. D. Wood, D. W. Parsons, S. Jones, T. Sjöblom, B. H. Park, R. Parsons, J. Willis, D. Dawson, J. K. Willson, T. Nikolskaya, Y. Nikolsky,

- L. Kopelovich, N. Papadopoulos, L. A. Pennacchio, T. L. Wang, S. D. Markowitz, G. Parmigiani, K. W. Kinzler, B. Vogelstein, V. E. Velculescu, Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 16224–16229 (2008).
11. D. Y. Chiang, G. Getz, D. B. Jaffe, M. J. O'Kelly, X. Zhao, S. L. Carter, C. Russ, C. Nusbaum, M. Meyerson, E. S. Lander, High-resolution mapping of copy-number alterations with massively parallel sequencing. *Nat. Methods* **6**, 99–103 (2009).
 12. J. O. Korbel, A. E. Urban, J. P. Affourtit, B. Godwin, F. Grubert, J. F. Simons, P. M. Kim, D. Palejev, N. J. Carriero, L. Du, B. E. Taillon, Z. Chen, A. Tanzer, A. C. Saunders, J. Chi, F. Yang, N. P. Carter, M. E. Hurler, S. M. Weissman, T. T. Harkins, M. B. Gerstein, M. Egholm, M. Snyder, Paired-end mapping reveals extensive structural variation in the human genome. *Science* **318**, 420–426 (2007).
 13. P. J. Campbell, P. J. Stephens, E. D. Pleasance, S. O'Meara, H. Li, T. Santarius, L. A. Stebbings, C. Leroy, S. Edkins, C. Hardy, J. W. Teague, A. Menzies, I. Goodhead, D. J. Turner, C. M. Cleve, M. A. Quail, A. Cox, C. Brown, R. Durbin, M. E. Hurler, P. A. Edwards, G. R. Bignell, M. R. Stratton, P. A. Futreal, Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat. Genet.* **40**, 722–729 (2008).
 14. C. A. Maher, C. Kumar-Sinha, X. Cao, S. Kalyana-Sundaram, B. Han, X. Jing, L. Sam, T. Barrette, N. Palanisamy, A. M. Chinnaiyan, Transcriptome sequencing to detect gene fusions in cancer. *Nature* **458**, 97–101 (2009).
 15. K. J. McKernan, H. E. Peckham, G. L. Costa, S. F. McLaughlin, Y. Fu, E. F. Tsung, C. R. Clouser, C. Duncan, J. K. Ichikawa, C. C. Lee, Z. Zhang, S. S. Ranade, E. T. Dimalanta, F. C. Hyland, T. D. Sokolsky, L. Zhang, A. Sheridan, H. Fu, C. L. Hendrickson, B. Li, L. Kotler, J. R. Stuart, J. A. Malek, J. M. Manning, A. A. Antipova, D. S. Perez, M. P. Moore, K. C. Hayashibara, M. R. Lyons, R. E. Beaudoin, B. E. Coleman, M. W. Laptewicz, A. E. Sannicandro, M. D. Rhodes, R. K. Gottmukkal, S. Yang, V. Bafna, A. Bashir, A. MacBride, C. Alkan, J. M. Kidd, E. E. Eichler, M. G. Reese, F. M. De La Vega, A. P. Blanchard, Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res.* **19**, 1527–1541 (2009).
 16. D. Grimwade, J. V. Jovanovic, R. K. Hills, E. A. Nugent, Y. Patel, R. Flora, D. Diverio, K. Jones, H. Aslett, E. Batson, K. Rennie, R. Angell, R. E. Clark, E. Solomon, F. Lo-Coco, K. Wheatley, A. K. Burnett, Prospective minimal residual disease monitoring to predict relapse of acute promyelocytic leukemia and to direct pre-emptive arsenic trioxide therapy. *J. Clin. Oncol.* **27**, 3650–3658 (2009).
 17. M. Bregni, S. Siena, A. Neri, R. Bassan, T. Barbui, D. Delia, G. Bonadonna, R. Dalla Favera, A. M. Gianni, Minimal residual disease in acute lymphoblastic leukemia detected by immune selection and gene rearrangement analysis. *J. Clin. Oncol.* **7**, 338–343 (1989).
 18. V. H. van der Velden, E. R. Panzer-Grümayer, G. Cazzaniga, T. Flohr, R. Sutton, A. Schrauder, G. Basso, M. Schrappe, J. M. Wijkhuijs, M. Konrad, C. R. Bartram, G. Masera, A. Biondi, J. J. van Dongen, Optimization of PCR-based minimal residual disease diagnostics for childhood acute lymphoblastic leukemia in a multi-center setting. *Leukemia* **21**, 706–713 (2007).
 19. T. Lion, Minimal residual disease. *Curr. Opin. Hematol.* **6**, 406–411 (1999).
 20. T. Hughes, M. Deininger, A. Hochhaus, S. Branford, J. Radich, J. Kaeda, M. Baccarani, J. Cortes, N. C. Cross, B. J. Druker, J. Gabert, D. Grimwade, R. Hehlmann, S. Kamel-Reid, J. H. Lipton, J. Longtine, G. Martinelli, G. Saglio, S. Soverini, W. Stock, J. M. Goldman, Monitoring CML patients responding to treatment with tyrosine kinase inhibitors: Review and recommendations for harmonizing current methodology for detecting *BCR-ABL* transcripts and kinase domain mutations and for expressing results. *Blood* **108**, 28–37 (2006).
 21. D. Dressman, H. Yan, G. Traverso, K. W. Kinzler, B. Vogelstein, Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 8817–8822 (2003).
 22. J. Shendure, G. J. Porreca, N. B. Reppas, X. Lin, J. P. McCutcheon, A. M. Rosenbaum, M. D. Wang, K. Zhang, R. D. Mitra, G. M. Church, Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**, 1728–1732 (2005).
 23. G. R. Bignell, T. Santarius, J. C. Pole, A. P. Butler, J. Perry, E. Pleasance, C. Greenman, A. Menzies, S. Taylor, S. Edkins, P. Campbell, M. Quail, B. Plumb, L. Matthews, K. McLay, P. A. Edwards, J. Rogers, R. Wooster, P. A. Futreal, M. R. Stratton, Architectures of somatic genomic rearrangement in human cancer amplicons at sequence-level resolution. *Genome Res.* **17**, 1296–1303 (2007).
 24. P. J. Stephens, D. J. McBride, M. L. Lin, I. Varela, E. D. Pleasance, J. T. Simpson, L. A. Stebbings, C. Leroy, S. Edkins, L. J. Mudie, C. D. Greenman, M. Jia, C. Latimer, J. W. Teague, K. W. Lau, J. Burton, M. A. Quail, H. Swerdlow, C. Churcher, R. Natrajan, A. M. Sieuwerts, J. W. Martens, D. P. Silver, A. Langerød, H. E. Russnes, J. A. Foekens, J. S. Reis-Filho, L. van 't Veer, A. L. Richardson, A. L. Borresen-Dale, P. J. Campbell, P. A. Futreal, M. R. Stratton, Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* **462**, 1005–1010 (2009).
 25. M. Li, F. Diehl, D. Dressman, B. Vogelstein, K. W. Kinzler, BEAMing up for detection and quantification of rare sequence variants. *Nat. Methods* **3**, 95–97 (2006).
 26. T. Sjöblom, S. Jones, L. D. Wood, D. W. Parsons, J. Lin, T. D. Barber, D. Mandelker, R. J. Leary, J. Ptak, N. Silliman, S. Szabo, P. Buckhaults, C. Farrell, P. Meeh, S. D. Markowitz, J. Willis, D. Dawson, J. K. Willson, A. F. Gazdar, J. Hartigan, L. Wu, C. Liu, G. Parmigiani, B. H. Park, K. E. Bachman, N. Papadopoulos, B. Vogelstein, K. W. Kinzler, V. E. Velculescu, The consensus coding sequences of human breast and colorectal cancers. *Science* **314**, 268–274 (2006).
 27. SOLiD data format and file definitions guide; http://www3.appliedbiosystems.com/cms/groups/mcb_marketing/documents/generaldocuments/cms_058717.pdf.
 28. Primer3 version 0.4.0; <http://frodo.wi.mit.edu/primer3/>.
 29. F. Diehl, K. Schmidt, M. A. Choti, K. Romans, S. Goodman, M. Li, K. Thornton, N. Agrawal, L. Sokoll, S. A. Szabo, K. W. Kinzler, B. Vogelstein, L. A. Diaz Jr., Circulating mutant DNA to assess tumor dynamics. *Nat. Med.* **14**, 985–990 (2008).
 30. Acknowledgments: We thank H. Peckham for genome mapping analyses of SOLiD data. Funding: NIH grants CA121113, CA057345, and CA043460; Lustgarten Foundation; National Colorectal Cancer Research Alliance; and United Negro College Fund–Merck Fellowship. Author contributions: R.J.L. and V.E.V. designed the study, generated and analyzed the data, and wrote the manuscript; I.K., F.D., and K.S. prepared clinical samples; C.C., C.D., and C.L. generated and sequenced SOLiD libraries; K.M. and F.M.D.V. provided advice for study design and analysis; K.W.K., B.V., and L.A.D. provided advice for study design, provided clinical samples, and wrote the manuscript. Competing interests: Under a licensing agreement between Johns Hopkins University and Genzyme, B.V., V.E.V., and K.W.K. are entitled to a share of royalties received by the University on sales of products related to research described in this paper. The terms of these arrangements are managed by Johns Hopkins University in accordance with its conflict-of-interest policies.

Submitted 7 December 2009

Accepted 4 February 2010

Published 24 February 2010

10.1126/scitranslmed.3000702

Citation: R. J. Leary, I. Kinde, F. Diehl, K. Schmidt, C. Clouser, C. Duncan, A. Antipova, C. Lee, K. McKernan, F. M. De La Vega, K. W. Kinzler, B. Vogelstein, L. A. Diaz Jr., V. E. Velculescu, Development of personalized tumor biomarkers using massively parallel sequencing. *Sci. Transl. Med.* **2**, 20ra14 (2010).

Development of Personalized Tumor Biomarkers Using Massively Parallel Sequencing

Rebecca J. Leary, Isaac Kinde, Frank Diehl, Kerstin Schmidt, Chris Clouser, Cisilya Duncan, Alena Antipova, Clarence Lee, Kevin McKernan, Francisco M. De La Vega, Kenneth W. Kinzler, Bert Vogelstein, Luis A. Diaz Jr., and Victor E. Velculescu

Sci. Transl. Med., 2 (20), • DOI: 10.1126/scitranslmed.3000702

PARE Personalizes Cancer Genetics

A diagnosis of cancer shatters the view of the world in an individual's mind. A world that once moved as comfortably as the pace of one's own life suddenly moves all too quickly. The individual starts asking questions and searching for possible remedies, and soon learns about the shockingly slow pace of successful cancer research. In the case of solid tumors, conventional surgical excision blanketed with "one size fits all" drug treatments simply fails to be universally effective in the long term. Cancer research has begun to shift to a more focused, personal approach that involves tailoring therapies directly to the complexity inherent in each individual—an area that holds considerable promise. But the differences among individuals are not the only layer of complexity hindering effective treatment. The intrinsic differences that accumulate over the course of tumor progression among similar tumor types hold the key to unlocking a truly personal remedy, a barcode, to cancer.

Now, Leary *et al.* make use of a massively parallel sequencing technique—personalized analysis of rearranged ends (PARE)—to home in on the unique DNA rearrangements present in tumors that differ from the rearrangements present in nontumor DNA from a small subset of individuals. They provide evidence for a highly sensitive, reliable, and cost-effective method, a foundation from which the annotation of large numbers of such tumor signatures will yield a personal cancer code. In an arena that takes small steps, PARE offers a leap forward in the clinical management and treatment of solid tumors, revealing true biomarkers that enable monitoring of individual tumor progression, tailoring of response to therapeutic treatment, and identification of residual disease at a level previously undetectable by current methods.

View the article online

<https://www.science.org/doi/10.1126/scitranslmed.3000702>

Permissions

<https://www.science.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of service](#)

Science Translational Medicine (ISSN 1946-6242) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Translational Medicine* is a registered trademark of AAAS. Copyright © 2010, American Association for the Advancement of Science

Supplementary Materials for

Development of Personalized Tumor Biomarkers Using Massively Parallel Sequencing

Rebecca J. Leary, Isaac Kinde, Frank Diehl, Kerstin Schmidt, Chris Clouser, Cisilya Duncan, Alena Antipova, Clarence Lee, Kevin McKernan, Francisco M. De La Vega, Kenneth W. Kinzler, Bert Vogelstein, Luis A. Diaz Jr., Victor E. Velculescu*

*To whom correspondence should be addressed. E-mail: velculescu@jhmi.edu

Published 24 February 2010, *Sci. Transl. Med.* 2, 20ra14 (2010)
DOI: 10.1126/scitranslmed.3000702

This PDF file includes:

Table S1. Comparison of SOLiD sequencing, Illumina SNP arrays, and Digital Karyotyping for analysis of copy number alterations.

Table S2. Putative copy number alterations identified by SOLiD sequencing in Co84 that were not identified by Illumina SNP arrays or Digital Karyotyping.

Table S3. Confirmed somatic rearrangements in breast and colorectal cancer samples.

Fig. S1. Flow chart of approach used to identify rearranged sequences.

Fig. S2. Comparison of Digital Karyotyping, Illumina SNP array, and SOLiD sequencing results on chromosome 8.

Table S1. Comparison of SOLiD sequencing, Illumina SNP arrays, and Digital Karyotyping for analysis of copy number alterations

	Tumor Sample	Chr	Digital Karyotyping				Illumina SNP Arrays				SOLiD sequencing			
			Left Boundary	Right Boundary	Size (bp)	Tag Density Ratio*	Left Boundary	Right Boundary	Size (bp)	Log R Ratio*	Left Boundary	Right Boundary	Size (bp)	Tag Density Ratio*
Amplification	Co84C	6	41,273,307	43,008,812	1,735,506	9.1	41,419,345	42,485,546	1,066,202	1.9	41,418,000	42,537,000	1,119,001	16.4
Amplification	Co84C	8	127,618,526	128,009,287	390,762	19.2	127,621,008	127,995,012	374,005	2.7	127,617,000	128,010,000	393,001	150.0
Amplification	Co84C	8	128,750,189	128,857,861	107,673	8.3	128,750,181	128,848,183	98,003	2.0	128,748,000	128,859,000	111,001	43.1
Amplification	Co84C	8	129,473,672	129,667,129	193,458	13.8	129,472,209	129,677,099	204,891	3.4	129,471,000	129,678,000	207,001	116.6
Amplification	Co84C	11	34,337,207	35,266,401	929,195	33.0	34,359,268	35,265,359	906,092	3.0	34,338,000	35,268,000	930,001	91.2
Amplification	Co84C	13	109,096,557	109,553,930	457,374	9.2	109,108,212	109,557,712	449,501	2.3	109,107,000	109,557,000	450,001	33.6
Amplification	Co84C	15	88,545,070	89,258,106	713,037	26.2	88,561,995	89,253,599	691,605	3.6	88,542,000	88,953,000	411,001	93.2
											88,983,000	89,118,000	135,001	32.8
											89,133,000	89,166,000	33,001	84.8
											89,208,000	89,256,000	48,001	50.3
Amplification	Co84C	19	34,570,450	34,641,949	71,500	7.9	34,561,976	34,641,548	79,573	2.2	34,548,000	34,641,000	93,001	33.9
Amplification	Co84C	19	34,956,853	35,344,522	387,670	14.3	34,966,463	35,321,409	354,947	2.6	34,956,000	35,346,000	390,001	36.8
Amplification	Co84C	19	36,274,262	36,388,331	114,070	6.2	36,281,540	36,385,232	103,693	2.5	36,273,000	36,396,000	123,001	21.2
Amplification	Co84C	19	54,500,237	54,643,655	143,419	8.4	54,520,709	54,622,533	101,825	2.1	54,498,000	54,636,000	138,001	41.8

* Values for Tag Density Ratios and Log R Ratios represent observed maximum values for amplifications.

Table S2. Putative copy number alterations identified by SOLiD sequencing in Co84 that were not identified by Illumina SNP arrays or Digital Karyotyping

Alteration Type	Chromosome	Left Boundary	Right Boundary	Size (bp)	Tag Density
					Ratio *
Homozygous deletion	1	83,388,000	83,532,000	144,001	0.0
Amplification	1	151,188,000	151,194,000	6,001	11.2
Amplification	1	159,393,000	159,414,000	21,001	9.7
Amplification	1	172,101,000	172,107,000	6,001	18.1
Amplification	1	179,910,000	179,916,000	6,001	17.4
Amplification	1	200,238,000	200,256,000	18,001	9.6
Amplification	1	204,168,000	204,186,000	18,001	13.2
Homozygous deletion	4	9,804,000	9,813,000	9,001	0.0
Homozygous deletion	4	69,066,000	69,171,000	105,001	0.0
Homozygous deletion	4	147,138,000	147,147,000	9,001	0.0
Amplification	5	31,749,000	31,755,000	6,001	12.3
Homozygous deletion	5	114,279,000	114,288,000	9,001	0.0
Homozygous deletion	7	38,358,000	38,364,000	6,001	0.0
Amplification	8	145,698,000	145,725,000	27,001	11.5
Homozygous deletion	10	66,978,000	66,984,000	6,001	0.0
Homozygous deletion	13	108,681,000	108,687,000	6,001	0.0
Amplification	13	110,139,000	110,157,000	18,001	22.5
Homozygous deletion	16	54,357,000	54,378,000	21,001	0.0
Homozygous deletion	16	59,112,000	59,130,000	18,001	0.0
Amplification	17	76,467,000	76,482,000	15,001	17.8
Homozygous deletion	18	14,268,000	14,289,000	21,001	0.0
Amplification	19	50,271,000	50,277,000	6,001	9.3
Amplification	20	25,404,000	25,428,000	24,001	13.1
Homozygous deletion	X	49,050,000	49,059,000	9,001	0.0
Homozygous deletion	X	121,650,000	121,734,000	84,001	0.0

* Values for Tag Density Ratios represent observed maximum values for amplifications.

Table S3. Confirmed somatic rearrangements in breast and colorectal cancer samples*

Sample	Forward tag		Reverse tag		Type	Primer 1	Primer 2
	Chromosome	Position	Chromosome	Position			
B5C	3	52,638,626	3	52,573,088	AAC	AAGTTTTTCAAGCTTTACCTGAAGT	TATATTGGAAGAATAGAAATGAATGG
B5C	4	93,109,700	4	-93,105,085	BAC	AGCCAAGTGCAATTCTCCAG	GCACACTGTTTGCGGAATG
B5C	11	57,713,780	8	-48,889,516	C**	GCCACCTTTCTTTCTTTCTGA	AAGCTTTGTTGGTTGTTCTCA
B5C	18	19,141,985	20	-29,591,944	C**	TGGCTTTCAAAACCCACTG	TCCTTTCTGCCCATAGGG
B5C	22	48,743,603	2	-104,047,142	C**	TCATGGTTTATCCACGGTGT	CACACCGCATTCACACAAAC
B7C	1	-96,237,189	7	65,542,257	C**	TCAAAACAGAAAGCATTAGGC	CGCATCCAAAGTATTAATAGCAA
B7C	2	197,428,606	2	113,761,988	AAC	AACCTCTCCACCTCAAAATC	CCAAATTGCCTGCTTAAGAGAT
B7C	2	-32,084,286	3	185,241,029	C**	TGCTACCAATACTTCCCACTTG	TACCGTCTCCAGGCATGT
B7C	2	114,604,628	18	53,562,784	C**	GGAGAAAACCTGGTTATTTTAA	TCCCTCATCAGAGCAATCA
B7C	3	-115,579,348	3	-115,651,310	AAC	AAATTGGGAAGGATCATACTGAC	TCTGAACATGCCTGATCTCATC
B7C	4	785,983	4	733,804	AAC	CTGAACCTCTGGGCTGAA	TTGTAAAGTGATGCTACCTGTG
B7C	5	107,405,959	5	107,231,803	AAC	CCTGGCCCTTAGGTAAAGAT	TGAAGAATCCTCTAGTGATGGAA
B7C	5	38,284,430	10	-44,715,202	C**	TGCAGCTTTTCTGTCTTCA	CTGCCAGTCCAACTGGTG
B7C	6	106,401,376	6	90,853,847	AAC	TGCTGTTTCAAAATTCCTACAGTC	TGAAATTAGGACCTGGAGCAC
B7C	6	101,933,981	6	102,444,426	ABC	GCCAGGTAACATGCTCACTTT	GATGCAGGAAGTGCACAGCA
B7C	9	22,003,033	9	21,761,298	AAC	GGGCTAAGCTTAAGAGTCTGG	GCCATGTGCAAGTCAAGAAG
B7C	11	-6,436,033	11	-6,519,897	AAC	TCTGCCCGCATACTGGAC	TAAGGGCGATGTGAACAAGG
B7C	12	65,950,588	12	65,923,399	AAC	GCCCTATTTTCAGAGAAAGTGTA	AACATCTCTTCCTTTGAAGATCC
B7C	13	60,438,525	13	52,159,979	AAC	AATTTGCTCTCATCGTATTGTGT	AGCTGAATCAAAATTTCCAATG
B7C	X	31,583,118	X	31,179,704	AAC	CTGAATCTCTTCCAGCAAAAT	AATGGGTAAAGCAGTTTAGGG
Co108C	2	191,184,628	5	-104,930,827	C**	TAGCATGCACCACTTTAGGC	AAAGGTAAAGGACTGTTTTAAGTTG
Co108C	2	78,849,963	6	-13,299,323	C**	GGTTCTGGAGGGTTGGAGA	GTTAAGATCAACATTTTTGTTTCAAG
Co108C	2	-7,268,710	6	13,299,385	C**	TATGCCACCATCGCTTAGGT	TCCAGTGCAATAAAACCAA
Co108C	2	-141,266,018	13	96,916,170	C**	GGTGTCTCTCTCCATACCA	CGATCTATACACCACCCACA
Co108C	3	-60,400,269	3	-60,437,489	AAC	TGCTTTTAGTTTGGGTACGG	GCTGATTGTTTATACCCAGTGC
Co108C	3	-60,365,933	3	-60,498,861	AAC	ATCCTCGGACTGGACTGAGA	AACCCCATCTGAAGCTACC
Co108C	3	60,573,034	3	60,472,593	AAC	GGGTATCTCAAAGGGCAGA	GCTCTCAATTTGTGTGATTGG
Co108C	4	81,934,151	15	54,039,041	C**	TGTGTTCTCTCTCTTAAGCAT	GACTACAAATGGCCAGACTC
Co108C	6	-13,299,291	5	157,523,537	C**	ATCCCCACATTTCCCAACC	CCCAGCCATATGTTGGTTTA
Co108C	6	13,299,271	2	-20,956,947	C**	GTATTTGTTTCATGTTTGGTGT	TCAATGGGGAGAGAGAGAC
Co108C	13	34,581,537	10	67,756,452	C**	ACGTGTGATTGGGGTAGC	CCAGATGGCTGGGTTAAATAAA
Co84C	8	128,442,121	19	49,144,200	C**	AGCTAGGTGGAGAAATTTGTCG	GGCTTCTGTAGAGTGCACATGA
Co84C	11	34,790,251	13	109,267,462	C**	AAGGAGATTGGTTATTGTGGAAA	CTGCAGAACTGTCTCATTTCTT
Co84C	11	-34,405,644	15	88,736,701	C**	TGCTGAATCATTCTCCCAACT	TGGTGATTCCACTGAGGTGA
Co84C	15	-89,096,347	8	127,747,412	C**	GCATTCTAAAGATGAAGTCCCAT	GGAAACCGTTAGTGGAAAAGTC
Hx402x	8	96,971,644	4	156,043,548	C**	CAGGTGATATACCAAGAAAATTAGG	TTTGGGTTCAATTCTATTTGAAGA
Hx402x	5	-100,413,406	5	-137,521,052	AAC	AGTCAACGCCCTAGCATGG	TGGGCATGAGCAAGATATTC
Hx402x	8	-144,771,376	8	-144,787,051	AAC	AATCAGTTGGGTGACTGTG	GTGACAGGCTGGGTGTCC
Hx402x	14	-85,526,541	14	-85,560,400	AAC	TGAAGGTTGAGTTGCCAGTG	TGTATGAAACATTGTAGAGGCTGT
Hx403x	1	119,547,240	1	-119,550,445	BBC	AGGAGGAAAGCAACACATAGAG	GGTGATTTTCAATGCATATTTCA
Hx403x	5	-27,160,637	5	27,150,736	BBC	AATTACCACAACTCCCAGCAG	CAAAAGATTTCCAAATGCAGTG
Hx403x	11	66,674,459	11	66,662,814	AAC	TGAATCAGAAAGTCTGGCAGT	CACCTGAGAAATCAATGATATGAG
Hx403x	16	6,343,641	16	6,727,736	AAC	CCTAGCCCTTTGTTCCCTGT	TTGTGTACCTAGACATTCATCCAA
Hx403x	16	6,574,321	16	6,759,729	AAC	GCAGAGAACAGCAGAAAAGTTG	AGCCCAAGATCAAGCCACAGA
Hx403x	16	26,579,136	16	-26,582,595	BBC	TTCTCTTCTCTGCCCTCAGTG	TTGATGATTTAGAAACTCTAGCCTGT
Hx403x	17	34,622,352	17	-34,624,284	BBA	GGCTCCCCCTCTCATTTCC	CTGCTGACGTGCTGGTCTT

* A single representative mate pair is shown for each rearrangement. Forward and reverse tags and their genomic coordinates correspond to F3 and R3 SOLiD mate pair tags. The type of rearrangement corresponds to the categories described in http://www3.appliedbiosystems.com/cms/groups/mcb_marketing/documents/generaldocuments/cms_058717.pdf. AAC corresponds to mate pairs spanning deletions; codes starting with B denote incorrect strand orientation; codes containing a B at the middle position denote incorrect ordering; and C** corresponds to interchromosomal translocations. Primers 1 and 2 correspond to primers used for confirming tumor-specific rearranged sequences.

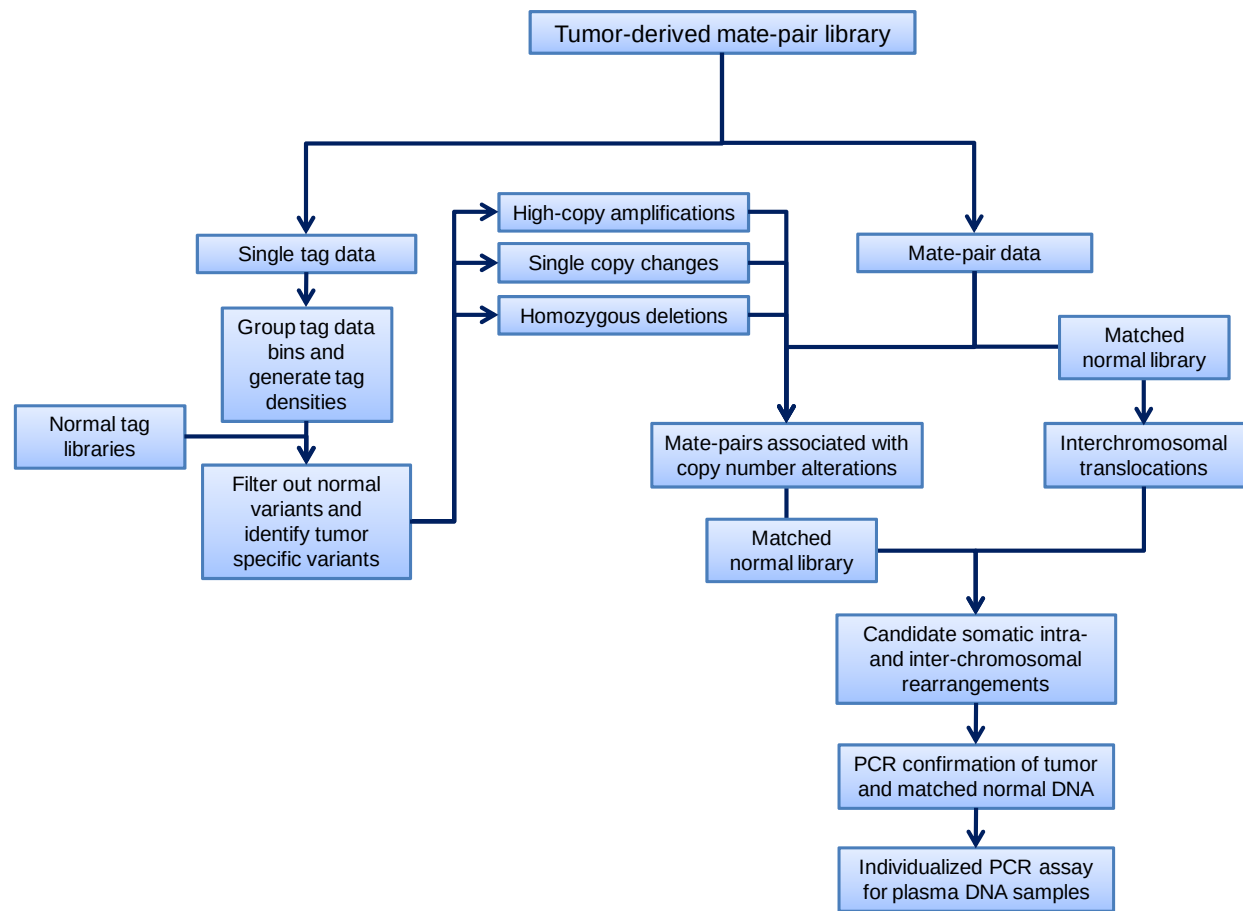


Fig. S1. Flow chart of approach used to identify rearranged sequences

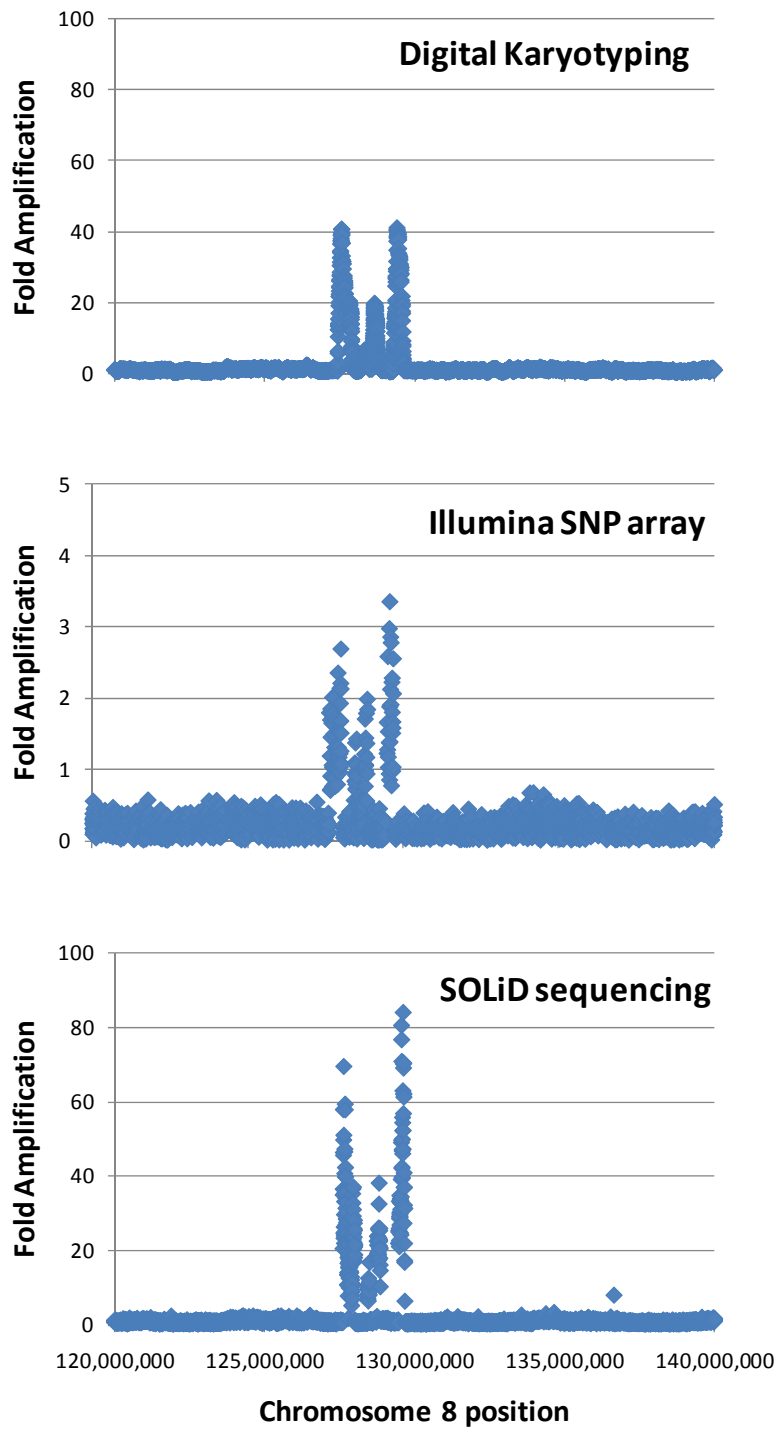


Fig. S2. Comparison of Digital Karyotyping, Illumina SNP array, and SOLiD sequencing results on chromosome 8.



Cancer Genome Scanning in Plasma: Detection of Tumor-Associated Copy Number Aberrations, Single-Nucleotide Variants, and Tumoral Heterogeneity by Massively Parallel Sequencing

K.C. Allen Chan,^{1,2,3} Peiyong Jiang,^{1,2} Yama W.L. Zheng,^{1,2} Gary J.W. Liao,^{1,2} Hao Sun,^{1,2} John Wong,⁴
Shing Shun N. Siu,⁵ Wing C. Chan,⁶ Stephen L. Chan,^{3,7} Anthony T.C. Chan,^{3,7} Paul B.S. Lai,⁴
Rossa W.K. Chiu,^{1,2} and Y.M.D. Lo^{1,2,3*}

BACKGROUND: Tumor-derived DNA can be found in the plasma of cancer patients. In this study, we explored the use of shotgun massively parallel sequencing (MPS) of plasma DNA from cancer patients to scan a cancer genome noninvasively.

METHODS: Four hepatocellular carcinoma patients and a patient with synchronous breast and ovarian cancers were recruited. DNA was extracted from the tumor tissues, and the preoperative and postoperative plasma samples of these patients were analyzed with shotgun MPS.

RESULTS: We achieved the genomewide profiling of copy number aberrations and point mutations in the plasma of the cancer patients. By detecting and quantifying the genomewide aggregated allelic loss and point mutations, we determined the fractional concentrations of tumor-derived DNA in plasma and correlated these values with tumor size and surgical treatment. We also demonstrated the potential utility of this approach for the analysis of complex oncologic scenarios by studying the patient with 2 synchronous cancers. Through the use of multiregional sequencing of tumoral tissues and shotgun sequencing of plasma DNA, we have shown that plasma DNA sequencing is a valuable approach for studying tumoral heterogeneity.

CONCLUSIONS: Shotgun DNA sequencing of plasma is a potentially powerful tool for cancer detection, monitoring, and research.

© 2012 American Association for Clinical Chemistry

The presence of tumor-derived DNA in the plasma of cancer patients offers exciting opportunities for the detection and monitoring of cancer (1, 2). Indeed, cancer-associated microsatellite alterations (3, 4), gene mutations (5–9), DNA-methylation changes (10, 11), and viral nucleic acids (12) have been found in the plasma of patients with different cancer types. Most of the previously published work on plasma DNA as a cancer marker has focused on the detection of specific and predetermined molecular targets known to be associated with cancer by means of such methods as the PCR (3, 4, 12), digital PCR (5–7), and digital ligation assays (9). With the advent of massively parallel sequencing (MPS),⁸ several groups have incorporated this approach for developing new plasma DNA-based cancer markers. One approach is to use MPS on tumor samples to first identify specific genomic rearrangements that can subsequently be detected in plasma (13, 14). Another approach is based on the use of targeted amplicon sequencing to search for mutations of genes that are commonly found in cancer (8).

¹ Li Ka Shing Institute of Health Sciences, ² Department of Chemical Pathology, ³ State Key Laboratory in Oncology in South China, Sir Y.K. Pao Centre for Cancer, ⁴ Department of Surgery, and ⁵ Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, New Territories, Hong Kong SAR, China; ⁶ Department of Surgery, North District Hospital, Sheung Shui, New Territories, Hong Kong SAR, China; ⁷ Department of Clinical Oncology, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, New Territories, Hong Kong SAR, China.

* Address correspondence to this author at: Department of Chemical Pathology, The Chinese University of Hong Kong, Prince of Wales Hospital, 30–32 Ngan

Shing St., Shatin, New Territories, Hong Kong SAR. Fax +852-26365090; e-mail loym@cuhk.edu.hk.

Received September 6, 2012; accepted September 27, 2012.

Previously published online at DOI: 10.1373/clinchem.2012.196014

⁸ Nonstandard abbreviations: MPS, massively parallel sequencing; HCC, hepatocellular carcinoma; SOAP2, Short Oligonucleotide Alignment Program 2; SNP, single-nucleotide polymorphism; LOH, loss of heterozygosity; LOESS, locally weighted scatterplot smoothing; SNV, single nucleotide variant; GAAL, genomewide aggregated allelic loss.

Owing to their targeted nature, the approaches outlined above can provide only a partial glimpse of the tumor genome in the plasma of cancer patients. For a genomewide view of the tumor genome in the circulation, a nontargeted random—or shotgun—sequencing approach would be desirable. In this regard, there has been much progress in the field of noninvasive prenatal diagnosis because of the results obtained with shotgun MPS of DNA from the plasma of pregnant women (15). This approach has allowed the noninvasive detection of fetal chromosomal aneuploidies (16–18) and fetal genomic scanning (19, 20).

In this article, we report the use of shotgun MPS to obtain a noninvasive, genomewide view of cancer-associated copy number variations and mutations in DNA in plasma. We have also sought to demonstrate the use of this approach for elucidating important tumoral characteristics, with tumoral heterogeneity used as an example.

Materials and Methods

SAMPLE COLLECTION

Hepatocellular carcinoma (HCC) patients and carriers of chronic hepatitis B were recruited from the Department of Surgery and the Department of Medicine and Therapeutics, respectively, of the Prince of Wales Hospital, Hong Kong, and informed consent and institutional review board approval were obtained. All HCC patients had Barcelona Clinic Liver Cancer stage A1 disease. Informed consent was obtained after the nature and possible consequences of the studies were explained. The patient with synchronous breast and ovarian cancers was recruited from the Department of Clinical Oncology, Prince of Wales Hospital. Peripheral blood samples from all participants were collected into EDTA-containing tubes. The tumor tissues of the HCC patients were obtained during their cancer-resection surgeries.

PROCESSING OF BLOOD

Peripheral blood samples were centrifuged at 1600g for 10 min at 4 °C. The plasma portion was recentrifuged at 16 000g for 10 min at 4 °C and then stored at –80 °C. Cell-free DNA molecules from 4.8 mL of plasma were extracted according to the blood and body fluid protocol of the QIAamp DSP DNA Blood Mini Kit (Qiagen). The plasma DNA was concentrated with a SpeedVac® Concentrator (Savant DNA120; Thermo Scientific) into a 40-μL final volume per case for subsequent preparation of the DNA-sequencing library.

GENOMIC DNA EXTRACTION

Genomic DNA was extracted from patients' buffy coat samples according to the blood and body fluid protocol

of the QIAamp DSP DNA Blood Mini Kit. DNA was extracted from tumor tissues with the QIAamp DNA Mini Kit (Qiagen).

DNA SEQUENCING

Sequencing libraries of the genomic DNA samples were constructed with the Paired-End Sample Preparation Kit (Illumina) according to the manufacturer's instructions. In brief, 1–5 μg genomic DNA was first sheared with a Covaris S220 Focused-ultrasonicator to 200-bp fragments. Afterward, DNA molecules were end-repaired with T₄ DNA polymerase and Klenow polymerase; T₄ polynucleotide kinase was then used to phosphorylate the 5' ends. A 3' overhang was created with a 3'-to-5' exonuclease-deficient Klenow fragment. Illumina adapter oligonucleotides were ligated to the sticky ends. The adapter-ligated DNA was enriched with a 12-cycle PCR. Because the plasma DNA molecules were short fragments (21) and the amounts of total DNA in the plasma samples were relatively small, we omitted the fragmentation steps and used a 15-cycle PCR when constructing the DNA libraries from the plasma samples.

An Agilent 2100 Bioanalyzer (Agilent Technologies) was used to check the quality and size of the adapter-ligated DNA libraries. DNA libraries were then measured by a KAPA Library Quantification Kit (Kapa Biosystems) according to the manufacturer's instructions.

The DNA library was diluted and hybridized to the paired-end sequencing flow cells. DNA clusters were generated on a cBot cluster generation system (Illumina) with the TruSeq PE Cluster Generation Kit v2 (Illumina), followed by 51 × 2 cycles or 76 × 2 cycles of sequencing on a HiSeq 2000 system (Illumina) with the TruSeq SBS Kit v2 (Illumina).

SEQUENCE ALIGNMENT AND FILTERING

The paired-end sequencing data were analyzed by means of the Short Oligonucleotide Alignment Program 2 (SOAP2) in the paired-end mode (22). For each paired-end read, 50 bp or 75 bp from each end was aligned to the non-repeat-masked reference human genome (Hg18). Up to 2 nucleotide mismatches were allowed for the alignment of each end. The genomic coordinates of these potential alignments for the 2 ends were then analyzed to determine whether any combination would allow the 2 ends to be aligned to the same chromosome with the correct orientation, spanning an insert size ≤600 bp, and mapping to a single location in the reference human genome. Duplicated reads were defined as paired-end reads in which the insert DNA molecule showed identical start and end locations in the human genome; the duplicate reads were removed as previously described (19).

MICROARRAY ANALYSIS

DNA extracted from the buffy coat and the tumor tissues of the HCC patients was genotyped with the Affymetrix Genome-Wide Human SNP Array 6.0 system, as previously described (23). The microarray data were processed with the Affymetrix Genotyping Console version 4.1. Genotyping analysis and single-nucleotide polymorphism (SNP) calling were performed with the Birdseed v2 algorithm, as previously described (24). The genotyping data for the buffy coat and the tumor tissues were used for identifying loss-of-heterozygosity (LOH) regions and for performing the copy number analysis. Copy number analysis was performed with the Genotyping Console with default parameters from Affymetrix and with a minimum genomic-segment size of 100 bp and a minimum of 5 genetic markers within the segment. Regions with LOH were identified as regions having 1 copy in the tumor tissue and 2 copies in the buffy coat, with the SNPs within these regions being heterozygous in the buffy coat but homozygous in the tumor tissue. For a genomic region exhibiting LOH in a tumor tissue, the SNP alleles that were present in the buffy coat but were absent from or of reduced intensity in the tumor tissues were considered to be the alleles on the deleted segment of the chromosomal region. The alleles that were present in both the buffy coat and the tumor tissue were deemed as having been derived from the nondeleted segment of the chromosomal region.

ARRAY COMPARATIVE GENOMIC HYBRIDIZATION ANALYSIS

DNA samples extracted from the buffy coat and the tumor tissues of the HCC patients were analyzed with the SurePrint G3 Human High Resolution Microarray Kit (Agilent) as previously described (25). Array comparative genomic hybridization data for the HCC patients were analyzed for copy number variation with the Partek® Genomics Suite. In brief, the raw probe intensities were adjusted according to the GC content of the sequence. This adjustment was followed by probe-level normalization of signal intensity while simultaneously adjusting for fragment length and probe sequences across all samples. Copy number gains and losses were detected by applying the default parameters of the Genomic Segmentation algorithm available in Partek Genomics Suite version 6.5 to obtain the different partitions of the copy number state.

DETECTION OF COPY NUMBER ABERRATION IN TUMOR TISSUE SAMPLES BY SEQUENCING

To investigate genomic copy number aberrations (e.g., copy number gains and copy number losses), we divided the genome into equal-sized segments (1 Mb per window/bin), and tallied the numbers of sequence reads mapping to each bin. Owing to the presence

of GC-dependent sequencing biases with high-throughput sequencing technologies (26), we used a statistical correction method, locally weighted scatterplot smoothing (LOESS), to correct the GC-associated bias (27). In this method, a correction factor is calculated for each bin according to the LOESS regression model, as previously described (28). Then, the read counts of each bin were adjusted with the bin-specific correction factor and normalized with the median read counts of all bins. After GC correction, a ratio of the adjusted read counts of the tumor to those of the buffy coat was calculated with the following equation:

$$R = \frac{A_{\text{tumor}}}{A_{\text{BC}}},$$

where A_{tumor} is the normalized GC-adjusted read counts of the tumor tissue and A_{BC} is the normalized GC-adjusted read counts of the buffy coat.

We then constructed a frequency distribution of $\log_2(R)$ for all bins. This distribution plot was used to estimate the proportion of tumor cells (F) in the tumor tissue that showed a particular copy number distribution and, subsequently, the copy number change at each bin.

On the frequency-distribution plot, a central peak at which R is approximately equal to 1 [i.e., $\log_2(R) = 0$] was identified; this peak represents the genomic regions without copy number aberrations. Then, the peaks lying to the left and right of the central peak were identified. These peaks represented regions with a 1-copy loss and a 1-copy gain, respectively. The distances of the left and right peaks from the central peak were used to determine the proportion of tumor cells (F) in the tumor tissue, according to the following equation:

$$F = R_{\text{right}} - R_{\text{left}},$$

where R_{right} is the R value of the right peak and R_{left} is the R value of the left peak.

Then, the copy number change (CN) values for all 1-Mb bins were calculated with the following equation:

$$\text{CN} = \frac{R_{\text{bin}} - R_{\text{cen}}}{0.5 \times F},$$

where R_{bin} is the R value of the bin and R_{cen} is the R value of the central peak.

DETECTION OF COPY NUMBER ABERRATIONS IN PLASMA

We analyzed the genomic representation of plasma DNA for different genomic regions. First, the entire genome was divided into 1-Mb windows, similar to the analysis of copy number aberrations in the tumor tissues. The GC-corrected read count was then determined (as described above) for each 1-Mb window. A z

score statistic was used to determine if the plasma DNA representation in a 1-Mb window would be significantly increased or decreased when compared with the reference group. The reference group consisted of the plasma samples from 16 healthy control individuals. In the current study, the GC-corrected read counts of each 1-Mb bin were normalized to the median GC-corrected read counts of all bins in the sample. The normalized plasma DNA representation was then compared with the data from the controls. A *z* score was then calculated for each 1-Mb window by using the mean and SDs of the controls. Regions with *z* scores of <-3 and >3 were regarded as significantly under- and overrepresented, respectively.

NUMBER OF MOLECULES REQUIRED FOR IDENTIFYING COPY NUMBER ABERRATIONS IN PLASMA

For copy number aberration analysis, the sensitivity and specificity of detecting tumor-associated copy number aberrations in plasma were determined by the precision of measuring the representation of plasma DNA in a chromosomal region and the fractional concentration of the tumor-derived DNA in the plasma of the cancer patient. The precision of measuring the plasma DNA representation in turn was affected by the number of plasma DNA molecules analyzed. In this regard, we performed simulation analyses to determine the relationship between the number of plasma DNA molecules required for analysis and the fractional concentration of tumor-derived DNA in the plasma so we could achieve a sensitivity of 95% for the detection of tumor-associated copy number aberrations. Computer simulations were performed for scenarios in which the affected region had a copy number change of -1 , $+1$, and $+2$ and for fractional concentrations of tumor-derived DNA ranging from 1% to 50%. In each simulation analysis, the entire genome was divided into 3000 bins. This number was similar to the one we used in the actual experimental analysis when a 1-Mb resolution was used.

We assumed that 10% of the bins would exhibit chromosomal aberrations in the tumor tissue. In the tumor tissue, the expected fraction (*P*) of total molecules falling into a bin within an affected region would be:

$$P = \frac{2 + \text{CN}}{2 \times 3000 + 3000 \times 10\% \times \text{CN}}$$

where CN is the copy number change. From this information, we calculated the expected change in the plasma.

In the plasma, the expected proportion of the total molecules (*E*) falling into a bin within an affected region can be calculated as:

$$E = P \times f + \frac{1 - f}{3000},$$

where *f* is the fractional concentration of tumor-derived DNA in plasma.

Simulations of 1000 normal cases and 1000 cancer cases were performed on the assumption of a binomial distribution of the plasma DNA molecules, with the expected plasma representations as calculated above and with an increasing number of molecules being analyzed until the 95% detection rate was reached. The simulation was conducted with the *rbinom* function in R (<http://www.r-project.org/>).

DETECTION OF TUMOR-ASSOCIATED SINGLE-NUCLEOTIDE VARIANTS

We sequenced the paired tumor and constitutional DNA samples to identify the tumor-associated single-nucleotide variants (SNVs). We focused on the SNVs occurring at homozygous sites in the constitutional DNA (i.e., buffy coat DNA). In principle, any nucleotide variation detected in the sequencing data of the tumor tissues but absent in the constitutional DNA could be a potential mutation (i.e., a SNV). Because of sequencing errors (0.1%–0.3% of sequenced nucleotides) (29), however, millions of false positives would be identified in the genome if a single occurrence of any nucleotide change in the sequencing data of the tumor tissue were to be regarded as a tumor-associated SNV. One way to reduce the number of false positives would be to institute the criterion of observing multiple occurrences of the same nucleotide change in the sequencing data in the tumor tissue before a tumor-associated SNV would be called. Because the occurrence of sequencing errors is a stochastic process, the number of false positives due to sequencing errors would decrease exponentially with the increasing number of occurrences required for an observed SNV to be qualified as a tumor-associated SNV. On the other hand, the number of false positives would increase exponentially with increasing sequencing depth. These relationships could be predicted with Poisson and binomial distribution functions. In this regard, we have developed a mathematical algorithm to determine the dynamic threshold of occurrence for qualifying an observed SNV as tumor associated. This algorithm takes into account the actual coverage of the particular nucleotide in the tumor sequencing data, the sequencing error rate, the maximum false-positive rate allowed, and the desired sensitivity for mutation detection.

In this study, we set very stringent criteria to reduce false positives. We required a mutation to be completely absent in the constitutional DNA sequencing, and the sequencing depth for the particular nucleotide position had to be >20 -fold. This threshold of occur-

rence was required to control the false-positive detection rate at $<1 \times 10^{-7}$. In this algorithm we also filtered out SNVs that were within centromeric, telomeric, and low-complexity regions to minimize false positives due to alignment artifacts. In addition, putative SNVs mapping to known SNPs in the dbSNP build 135 database were also removed.

Results

TUMOR-ASSOCIATED COPY NUMBER ABERRATIONS IN PLASMA

We investigated whether tumor-associated copy number aberrations could be detected in the plasma of cancer patients by shotgun MPS. Peripheral blood samples were obtained both before and 1 week after surgical resection with curative intent from 4 HCC patients. The blood samples were fractionated into plasma and blood cells. DNA was also obtained from each of the tumors. Copy number aberrations in the 4 tumor samples were analyzed with MPS and with 1 or 2 microarray platforms (Affymetrix and Agilent). Copy number aberrations were analyzed in 1-Mb windows across the genome in the tumor tissues and compared with the plasma samples from a group of 16 healthy control individuals. The data were consistent across the 3 platforms (see Fig. 1 in the Data Supplement that accompanies the online version of this article at <http://www.clinchem.org/content/vol59/issue1>).

We then used MPS to analyze the pre- and post-resection plasma samples obtained from all 4 HCC patients. The mean sequencing depth was 17-fold coverage of the haploid human genome (range, 15.2-fold to 18.5-fold). Fig. 1 shows Circos plots (30) of the copy number aberrations across the genome in the tumor, the pre-resection plasma sample, and the post-resection plasma sample, for each patient. In each case, characteristic copy number aberrations seen in the tumor tissue sample were also observed in the pre-resection plasma sample (Fig. 1). A significant change in the regional representation of plasma DNA was defined as >3 SDs from the mean representation of the 16 healthy controls for the corresponding 1-Mb window.

For all cases, such copy number aberrations disappeared almost completely in the post-resection plasma sample (Fig. 1). The detectability of the different classes of tumor-associated genetic alterations in plasma is shown in Fig. 2. For comparison, we used the same approach to analyze plasma DNA samples from 4 hepatitis B carriers without HCC (Fig. 1E; see Fig. 2 in the online Data Supplement). These individuals were followed up for 1 additional year after blood sampling and had no evidence of HCC. For these individuals, 99% of the sequenced bins showed normal representations in plasma (see Table 1 in the online Data Supplement).

Similarly, a mean of 98.9% of the sequenced bins in the 16 healthy controls showed normal representations in plasma (see Table 2 and Fig. 3 in the online Data Supplement). These results indicate that the analysis of copy number aberrations in plasma is specific for differentiating between cancer patients and individuals without a cancer; however, the specificity for plasma copy number analysis appeared to be reduced in the HCC patients. Hence, in the 4 HCC patients, a median of 15% (range, 2%–48%) of the regions at which no copy number aberrations occurred in the corresponding tumor tissue showed an aberrant plasma DNA representation (see Table 1 in the online Data Supplement). This issue will be discussed in more detail in the Discussion section.

FRACTIONAL CONCENTRATION OF TUMOR DNA IN PLASMA DETERMINED BY GENOMEWIDE AGGREGATED ALLELIC LOSS ANALYSIS

The fractional concentrations of tumor-derived DNA in plasma were determined by analyzing, in a genome-wide manner, the allelic counts for SNPs exhibiting LOH in the plasma shotgun MPS data, which we term “genomewide aggregated allelic loss” (GAAL) analysis. For such an analysis, we chose SNPs that exhibited LOH in the tumors as demonstrated with the Affymetrix SNP 6.0 microarray. The alleles deleted in the tumors would have lower concentrations in the plasma than those that were not deleted. The difference in their concentrations was related to the concentration of tumor-derived DNA in the plasma sample. Thus, the plasma concentration of the tumor-derived DNA (C) can be deduced with the following equation:

$$C = \frac{N_{\text{nondel}} - N_{\text{del}}}{N_{\text{nondel}}},$$

where N_{nondel} represents the number of sequenced reads carrying the nondeleted alleles in the tumor tissues, and N_{del} represents the number of sequenced reads carrying the deleted alleles in the tumor tissues.

Table 1 lists the fractional concentrations of tumor DNA in the plasma samples for each of the 4 cases. The size of the tumor appears to be correlated with the estimated fractional concentration of tumor-derived DNA in plasma before surgical resection. For example, we estimated that tumor-derived DNA accounted for 52% of the total plasma DNA in the patient who had the largest tumor (13 cm) of the 4 HCC cases. For each of the 4 cases, we observed a reduction in the fractional concentration of tumor-derived DNA after surgical resection of the tumor (Table 1).

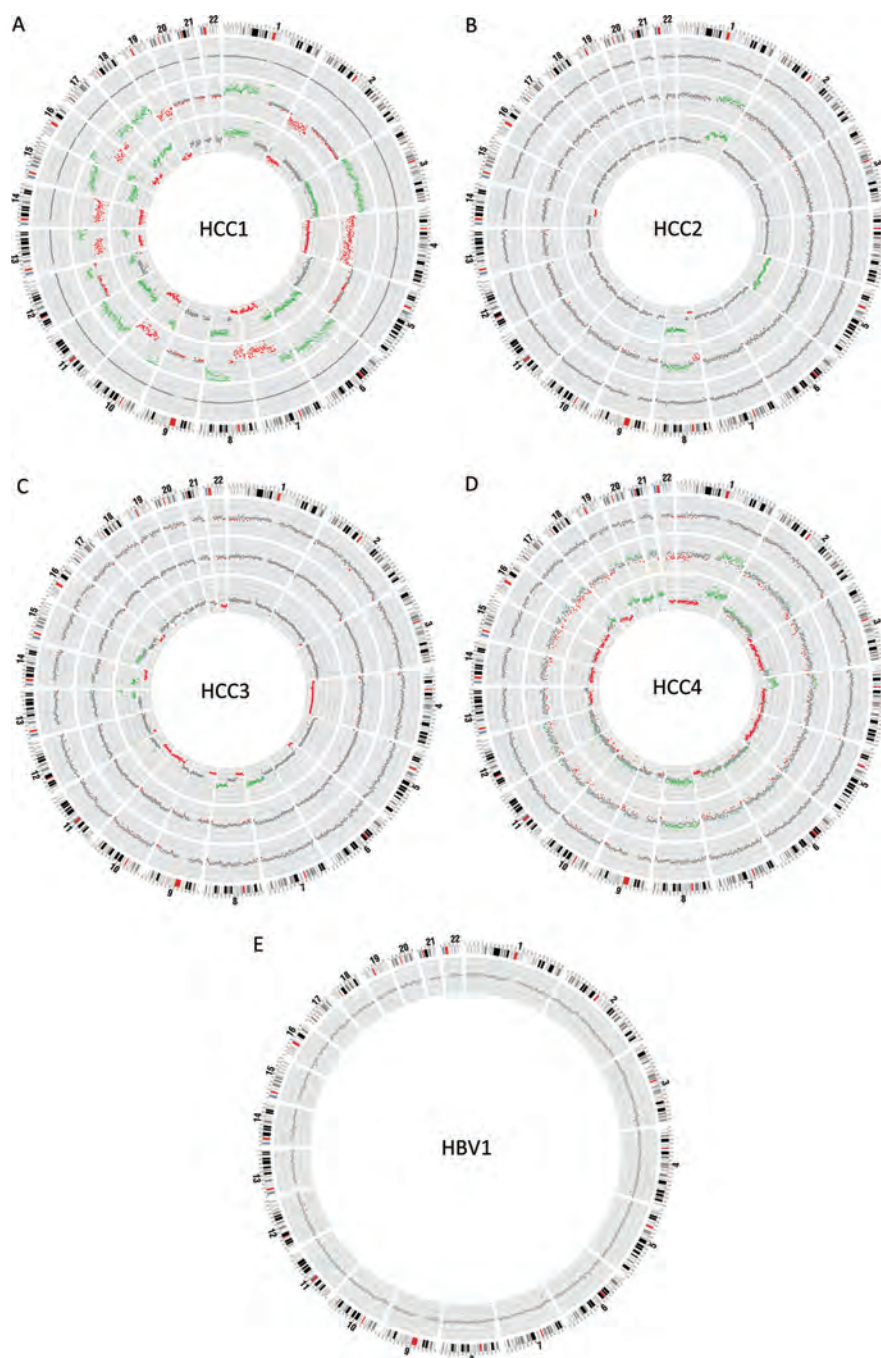
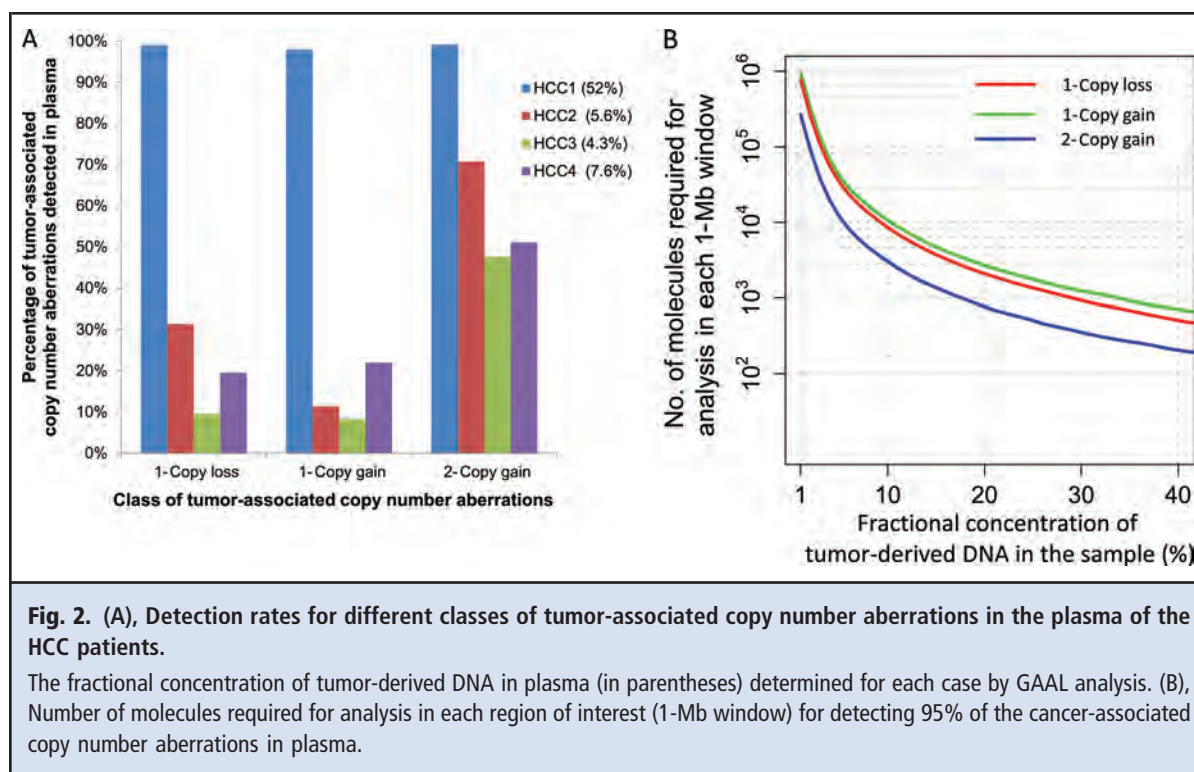


Fig. 1. Copy number aberrations detected in the tumor tissue sample (inner ring), presurgery plasma sample (middle ring), and postsurgery plasma sample (outer ring) for the 4 HCC cases (A–D) and genomic representation of plasma DNA samples for a hepatitis B virus carrier (HBV1) without HCC (E).

Chromosome ideograms (outside the plots) are oriented pter to qter in a clockwise direction. For the tumor tissues, the numbers of copies gained (green) or lost (red) are presented. For the plasma samples, regions with increased and reduced representation in plasma are shown in green and red, respectively. Regions with z scores between -3 and $+3$ are represented by gray dots. For HCC1, the distance between 2 consecutive horizontal lines represents a z score of 30. For the other cases, the distance represents a z score of 10.



FACTORS INFLUENCING THE DETECTION OF COPY NUMBER ABERRATIONS IN PLASMA

The fractional concentration of tumor DNA in plasma and the class of copy number aberrations strongly influenced the detectability of such alterations in plasma. Fig. 1 and Fig. 2A show that the proportions of tumor-associated genetic aberrations that could be seen in the plasma DNA-sequencing results were correlated with the fractional concentrations of tumor DNA in plasma. For

example, case HCC1, which had the largest tumor and the highest fractional concentration of tumor DNA in plasma, also had the largest proportion of tumor-associated copy number aberrations detected in plasma (Fig. 1A).

Case HCC1 has a fractional concentration of tumor-derived DNA of 52%. Before treatment, most of the tumoral copy number aberrations could also be seen in the plasma. At most of the chromosomal re-

Table 1. Fractional concentration of tumor-derived DNA in plasma by GAAL analysis.

Case	Maximal dimension of the tumor, cm	SNPs with LOH in tumor, n	Plasma			
			Collection time	N_{nondel}^a	N_{del}	Fractional concentration of tumor-derived DNA
HCC1	13.0	11 310	Before surgery	88 499	42 829	52%
			After surgery	94 442	93 584	0.9%
HCC2	6.2	6040	Before surgery	56 611	53 428	5.6%
			After surgery	63 694	63 083	0.9%
HCC3	4.2	24 783	Before surgery	198 688	190 163	4.3%
			After surgery	236 655	233 280	1.4%
HCC4	6.2	498	Before surgery	2465	2277	7.6%
			After surgery	2773	2699	2.7%

^a N_{nondel} , number of sequenced reads carrying the nondeleted alleles in tumor tissues; N_{del} , number of sequenced reads carrying the deleted alleles in the tumor tissues.

Table 2. Fractional concentrations of tumor-derived DNA in plasma determined by SNV analysis.

Case	No. of SNVs detected in tumor tissue	Time point	No. of tumor-associated SNVs sequenced from plasma (percentage of SNVs seen in tumor tissue)	No. of plasma DNA sequence reads showing SNVs (<i>p</i>)	No. of plasma DNA sequence reads showing wild-type sequence (<i>q</i>)	Deduced fractional concentration of tumor-derived DNA by SNV analysis ($\frac{2p}{p+q}$)	Deduced fractional concentration of tumor-derived DNA by GAAL analysis
HCC1	2840	Before surgery	2569 (94%)	11 389	31 602	53%	52%
		After surgery	44 (1.5%)	91	46 898	0.4%	0.9%
HCC2	3105	Before surgery	1097 (35%)	1490	57 865	5.0%	5.6%
		After surgery	72 (2.3%)	206	66 692	0.6%	0.9%
HCC3	3171	Before surgery	461 (15%)	525	48 886	2.1%	4.3%
		After surgery	31 (1%)	67	58 862	0.2%	1.4%
HCC4	1334	Before surgery	201 (15%)	248	18 527	2.6%	7.6%
		After surgery	74 (5.5%)	149	22 144	1.3%	2.7%

gions with a single-copy gain in the tumor (e.g., chromosomes 1p, 3, and 6), the *z* scores were >20, indicating that the plasma representation was 20 SDs above the mean representation of the healthy control individuals for these regions. On the other hand, case HCC3 had a fractional tumor DNA concentration in plasma of 4.3%, and a smaller proportion of the cancer-associated aberrations could be observed in the plasma. None of the regions with a single-copy gain (7q, 8q, 13q, and 14p) had a *z* score of >10 in the plasma.

The classes of copy number aberrations that were studied included 1-copy losses, 1-copy gains, and 2-copy gains. The percentages of such changes that were detected in plasma in each of the 4 HCC cases are plotted in Fig. 2A and listed in Table 3 in the online Data Supplement. For each case, 2-copy gains could be detected with higher sensitivity in plasma than 1-copy changes. For all 4 HCC patients, most of these tumor-associated chromosomal aberrations disappeared after surgical resection of the tumor (Fig. 1; see Table 3 in the online Data Supplement).

Previous work on noninvasive prenatal diagnosis has revealed that a greater sequencing depth would enable the detection in plasma of an aneuploid fetus at a lower fractional fetal DNA concentration (31). Using computer simulation, we explored the relationship between the depths of sequencing that would be needed to detect different classes of tumor-associated copy number aberrations in plasma at different fractional concentrations of tumor DNA in plasma (Fig. 2B). For illustration purposes we fixed the detection rate at 95% and explored 3 classes of genetic aberrations—1-copy loss, 1-copy gain, and 2-copy gain. When the fractional concentration of tumor-derived DNA was 40%, the detection of aberrations with 2-copy gains and 1-copy

gains would require the analysis of approximately 180 and 800 molecules, respectively, per 1-Mb window. When the fractional concentration of tumor-derived DNA drops to 10%, the analysis of approximately 2500 and 12 000 molecules per 1-Mb window is necessary to detect these respective changes. The requirement of an exponential increase in the number of molecules with a decreasing fractional concentration of tumor-derived DNA was consistent with the requirement for noninvasive prenatal diagnosis of fetal chromosomal aneuploidies via analysis of maternal plasma DNA (32).

TUMOR-DERIVED SNVs IN PLASMA

We next explored the genomewide detection of tumor-derived SNVs in the plasma of the 4 HCC patients. We sequenced tumor DNA and buffy coat DNA to mean depths of 29.5-fold (range, 27-fold to 33-fold) and 43-fold (range, 39-fold to 46-fold) haploid genome coverage, respectively. The MPS data from the tumor DNA and the buffy coat DNA from each of the 4 HCC patients were compared, and SNVs present in the tumor DNA but not in the buffy coat DNA were mined with a stringent bioinformatics algorithm. This algorithm required a putative SNV to be present in at least a threshold number of sequenced tumor DNA fragments before it would be classified as a true SNV. The threshold number was determined by taking into account the sequencing depth of a particular nucleotide and the sequencing error rate.

The number of tumor-associated SNVs ranged from 1334 to 3171 in the 4 HCC cases. The proportions of such SNVs that were detectable in plasma are listed in Table 2. Before treatment, 15%–94% of the tumor-associated SNVs were detected in plasma. The fractional concentrations of tumor-derived DNA in plasma were

determined by the fractional counts of the mutant with respect to the total (i.e., mutant plus wild type) sequences (Table 2). These fractional concentrations were well correlated with those determined with GAAL analysis and were reduced after surgery (Table 2).

To estimate the specificity of the SNV analysis approach, we analyzed the plasma of the healthy controls for the tumor-associated SNVs (see Table 4 in the online Data Supplement). The presence of a small number of these putative tumor-associated mutations in the plasma of the 16 healthy controls represented the “stochastic noise” of this method and was likely due to sequencing errors. The mean fractional concentration estimated from such noise was 0.38%.

PLASMA DNA ANALYSIS IN A PATIENT WITH MULTIPLE SYNCHRONOUS CANCERS

To illustrate the use of shotgun MPS of plasma DNA to monitor a complex oncologic scenario, we studied a 58-year-old female patient with a *BRCA1* (breast cancer 1, early onset) mutation (p.Cys1697*) presenting with synchronous breast and ovarian cancers (Fig. 3). The breast cancer was a 3-cm infiltrating ductal carcinoma located in the left breast. The patient had serous adenocarcinomas of both ovaries. The tumor in the left ovary measured 6 cm in the longest dimension, and the one on the right side measured 12 cm. There were also multiple intra-abdominal tumor deposits involving the omentum and the colon. Surgical resections of the breast tumor and the ovarian tumors, together with the omentum and the sigmoid colon, were performed on the same day. The breast tumor and the ovarian tumor tissues from the left and right sides were collected for this study. Plasma samples were also collected at diagnosis and 1 day after the operation. The breast tumor DNA and 4 regions of the ovarian tumors (2 from the left ovary and 2 from the right ovary) were analyzed by MPS. The regions sampled for the ovarian tumor on the same side were separated by 4 cm.

Copy number aberrations for the breast and ovarian cancers are plotted in Fig. 4A. The 4 ovarian tumoral regions showed highly similar patterns of copy number aberrations. On the other hand, the breast cancer exhibited a different pattern of copy number aberrations. The pattern of aberrations in the presurgery plasma sample was a composite of the patterns of aberrations in the breast and ovarian cancers (Fig. 4B). Examples of genetic aberrations that were specific to the breast cancer included a deleted segment on chromosome 6p and amplifications on chromosomes 1q, 7p, and 15q (Fig. 4A). On the other hand, examples of genetic aberrations that were specific to the ovarian tumors included deletions on chromosomes 2, 4p, 11p, 12q, 18q, and 22q

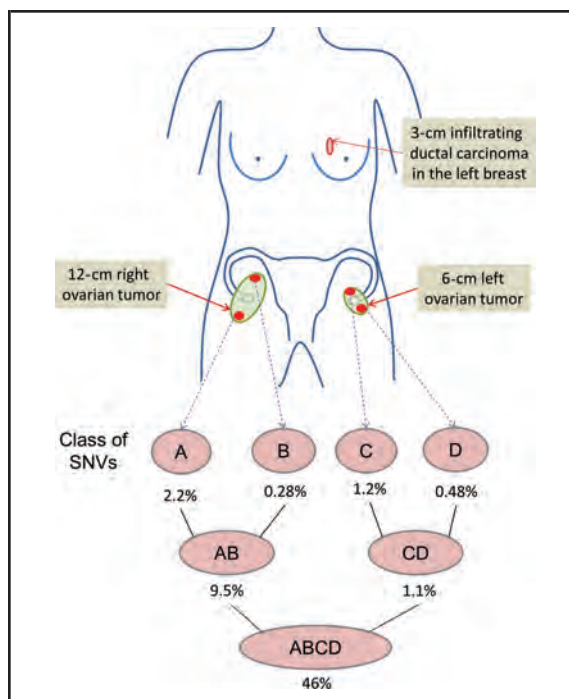


Fig. 3. Illustration showing the locations of the tumors in a patient with synchronous breast and ovarian cancers.

Four regions (A, B, C, and D) from the ovarian tumors were sampled. Shown are the fractional concentrations of tumor-derived plasma DNA estimated with SNVs associated with each of the ovarian tumoral regions. The calculated fractional concentration of tumor DNA in plasma increases when the estimate is based on SNVs associated with multiple tumor regions.

and amplifications on chromosomes 3q, 5p, and 21q (Fig. 4B). These genetic aberrations specific to the breast cancer or to the ovarian cancers were all observed in the presurgery plasma sample but were cleared in the postsurgery plasma sample (Fig. 4A). An expanded view of 2 genomic regions exhibiting copy number aberrations that were present in the breast cancer but absent in the ovarian tumors is shown in Fig. 4C.

To examine the relative contributions made by the breast and ovarian cancers to the plasma DNA of the patient, we conducted GAAL analyses for genomic regions that exhibited deletions specific to either the breast tumor or the ovarian tumors. These results indicated that the breast cancer and the ovarian cancers contributed 2.1% and 46%, respectively, of the DNA in plasma before surgery (see Table 5 in the online Data Supplement). The fractional concentrations of DNA contributed by each of these tumors dropped to 1.3%

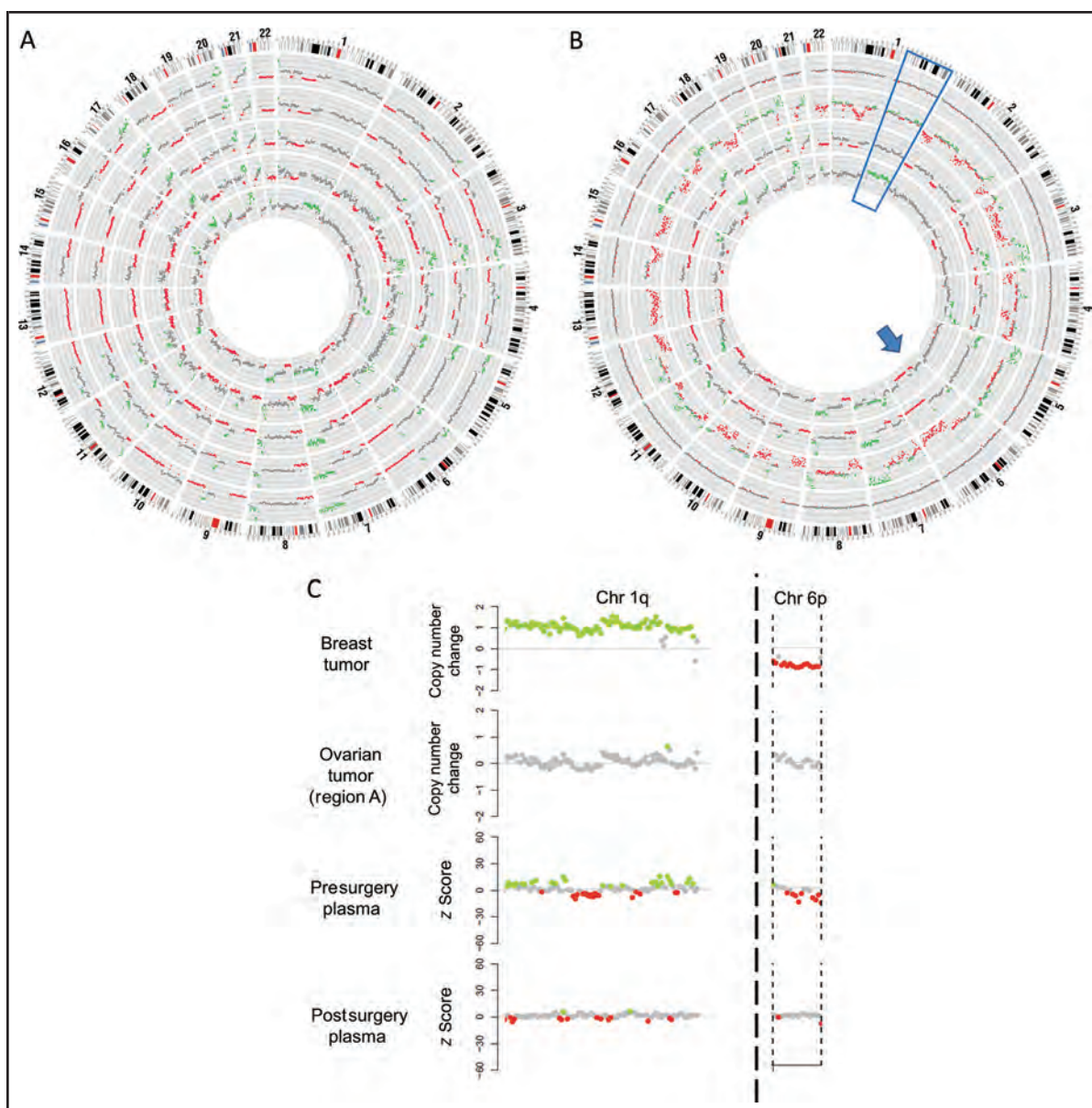


Fig. 4. (A), Copy number aberrations detected in the breast cancer (innermost ring) and the ovarian tumors (second innermost ring outwards: regions A, B, C, and D).

(B), Copy number aberrations in the breast tumor, ovarian tumor (region A), presurgery plasma, and postsurgery plasma (from inside to outside). Two genomic regions exhibiting copy number aberrations specific to the breast cancer are marked with a box (chromosome 1q) and an arrow (chromosome 6p). (C), Expanded view of the chromosome 1q and 6p regions in tumor and plasma samples.

and 0.66%, respectively, after surgery (see Table 5 in the online Data Supplement).

INVESTIGATION OF TUMORAL HETEROGENEITY

We further explored the phenomenon of tumoral heterogeneity (33) by studying the 4 regions sampled from

the ovarian tumors. There were no observable differences among these 4 regions in terms of copy number aberrations (Fig. 4A). We compared the SNV profile from each of these regions with the buffy coat DNA of the patient. We classified the SNVs into 7 groups: 4 groups containing mutations that were unique to each

Table 3. SNV analysis of multiple regions of the ovarian cancer in the patient with synchronous cancers.

Group of SNVs	Presurgery plasma				Postsurgery plasma		
	No. of SNVs in tumor	No. of sequence reads showing SNVs in plasma (p)	No. of sequence reads showing wild-type sequence in plasma (q)	Deduced fractional concentration of tumor-derived DNA $\left(\frac{2p}{p+q}\right)$	No. of sequence reads showing SNVs in plasma (p)	No. of sequence reads showing wild-type sequence in plasma (q)	Deduced fractional concentration of tumor-derived DNA $\left(\frac{2p}{p+q}\right)$
A	71	37	3321	2.2%	5	2149	0.46%
B	122	8	5633	0.28%	1	3516	0.06%
C	168	51	8507	1.2%	2	5438	0.07%
D	248	26	10 880	0.48%	2	7060	0.06%
AB	10	21	423	9.5%	0	297	0%
CD	12	3	569	1.1%	0	333	0%
ABCD	2216	21 344	70 940	46%	54	61 417	0.18%

region (i.e., groups A, B, C, and D), 2 groups that contained mutations shared by the 2 regions on each side of the body (i.e., groups AB and CD), and a final group containing mutations shared by all 4 regions (i.e., group ABCD) (Table 3 and Fig. 3). We randomly selected approximately 10 SNVs for each of the 7 SNV groups for validation with a mass spectrometry-based single-nucleotide extension method (iPLEX analysis; Sequenom) (see Table 6 in the online Data Supplement). A total of 67 mutations were subjected to validation. More than 95% of the SNV results determined by sequencing were validated in the iPLEX analyses.

SNVs from each of these groups were then sought out in the plasma shotgun MPS data. The fractional concentrations of circulating tumor DNA were determined with each SNV group (Table 3). The fractional concentrations of tumor DNA in plasma before surgery and after surgery, as determined by SNVs shared by all 4 regions (i.e., group ABCD), were 46% and 0.18%, respectively. These latter percentages correlated well with those obtained in GAAL analyses (see Table 5 in the online Data Supplement). The fractional concentrations of tumor-derived DNA in pre-operative plasma determined with SNVs from groups AB and CD were 9.5% and 1.1%, respectively (Table 3). These concentrations were consistent with the relative sizes of the right and left ovarian tumors (Fig. 3). The fractional concentrations of tumor-derived DNA determined with the region-unique SNVs (i.e., those in groups A, B, C, and D) were generally low. The trend of a reduction in the observed fractional tumor-derived DNA in plasma as measured with SNVs of increasing “regional specificity” will be explored in the Discussion.

Discussion

Our data indicate that shotgun sequencing of plasma samples from cancer patients would allow cancer-associated copy number aberrations and mutations to be analyzed noninvasively and in a genomewide fashion (Figs. 1 and 4). This finding is an important step forward from previous work, which has generally been focused on detecting a small number of tumor-associated genetic changes in plasma (1). This approach would allow genomic aberrations harbored by tumor cells within a patient to be scanned at different levels of resolution, ranging from a global view of cancer-associated copy number aberrations to point mutations carried by different clones of tumor cells. This approach would allow qualitative and quantitative aspects of such aberrations to be followed serially, as evidenced by the changes we observed in the pre- and postsurgery plasma samples. The method would also allow tumor load to be assessed, as indicated by the correlation between the fractional tumor DNA concentration measured and the size of the tumor. In this study, these general characteristics were observed in the plasma DNA profiles derived from HCC, breast cancer, and ovarian cancer.

Both GAAL analysis and SNV analysis allow the fractional concentrations of tumor-derived DNA in plasma to be measured for any cancer. These approaches would allow one to compare DNA released into the plasma from different types of tumors. Through the pooling of the sequence counts from SNPs involved in LOH and point mutations across the genome, the measurements made by these analyses are expected to be much more precise than those made on

the basis of individual tumor-associated genetic changes.

The control group for establishing the baseline for plasma-based copy number analysis consisted of healthy individuals without cancer. The specificity of this type of analysis for individuals without cancer can be seen from bootstrap analysis of this control group, as well as in chronic hepatitis B carriers without HCC. For these 2 groups, only a small fraction of regions had aberrant plasma DNA representation with a z score <-3 or >3 . These results indicate that this approach is specific for differentiating between cancer patients and individuals without cancer. When this approach was used for analyzing the plasma of HCC patients, however, the specificity appeared to be lower than for study participants without cancer. The explanation for this observation is not entirely clear at the moment; however, we can think of at least 2 possible explanations. First, the presence of tumor DNA with regions exhibiting copy number aberrations in the plasma of cancer patients would affect the observed relative contribution of DNA from regions not exhibiting copy number aberrations. For example, for a tumor exhibiting many copy number losses across the genome, one would observe a relatively increased contribution of plasma DNA originating from genomic regions with a normal copy number. This effect would be greatest for plasma samples containing high concentrations of tumor DNA. Indeed, the percentages of misclassification of normal regions are the highest for case HCC1 (see Table 1 in the online Data Supplement), which has the highest fractional tumor DNA concentration in plasma (52%). Second, the phenomenon of tumoral heterogeneity could explain a proportion of such apparent false-positive results. In other words, the single tumoral region sampled from each HCC tumor might not contain a particular copy number aberration that has been released into the plasma from another tumor clone not contained in the sampled region. Although the phenomenon mentioned in this paragraph requires more study, it does not appear at present to adversely affect the ability to distinguish between individuals with cancer and those without cancer.

The patient with synchronous breast and ovarian cancers is particularly noteworthy, because the case illustrates a number of important concepts. First, it shows that careful analysis of the sequencing data permits one to dissect out the presence of DNA in the plasma originating from individual tumor cell populations. Targeting genomic regions exhibiting copy number aberrations specific to the breast and ovarian cancers allowed the elucidation of the relative contributions of these tumors to the circulating DNA pool (Fig. 4B; see Table 5 in the online Data Supplement).

To study the effect of tumoral heterogeneity on the plasma DNA profile, we analyzed the mutational profile of the 4 regions of the bilateral ovarian cancer via comparisons with the patient's constitutional DNA. We were able to group these mutations into 7 categories according to the degree of sharing of these mutations between the 4 regions. It is interesting that mutations that were shared by all 4 regions (i.e., group ABCD) contributed the highest fractional contribution of tumor-derived DNA to the plasma. On the other hand, mutations that were more region specific had a reduced contribution to plasma. It was therefore not surprising that the GAAL analysis, which is based on aggregating the amount of allelic loss across the genome, produced a fractional tumor DNA concentration similar to that of the analysis based on SNVs from the ABCD category. These data suggest that for an accurate measurement of the total tumor load in a cancer patient, the use of a genomewide shotgun approach might provide a more representative picture, compared with the more traditional approach of targeting specific tumor-associated mutations. For the latter approach, if only a subset of the tumor cells possesses the targeted mutations, one might miss important information regarding imminent relapse or disease progression caused by tumor cells not possessing the targeted mutations, or one might miss the emergence of a treatment-resistant clone.

Indeed, the phenomenon of tumoral heterogeneity (33) has created challenges for developing truly representative tumor markers for monitoring purposes. Because plasma receives DNA from the various heterogeneous tumor clones in the body, shotgun sequencing of plasma DNA might be a readily available and noninvasive method for studying and monitoring tumoral heterogeneity and the total tumor burden in the body.

In this work, we used tumor-associated SNVs as tumor markers and did not focus on elucidating the biology of the mutational landscape of the studied cancers. Thus, the present work has not carried out a detailed characterization of the mutated sequences. We expect, however, that most of the tumor-associated SNVs we detected were "passenger" rather than "driver" mutations. It was the aim of this study to develop a noninvasive approach to assess the tumor burden of the body. We therefore believe that the detection of mutations—passengers and drivers—in a genomewide manner provides a numerical advantage that has the potential for more sensitive, precise, and representative detection of tumor DNA in plasma. On the other hand, selective analyses of driver mutations may allow more readily the identification of medically actionable targets (8). Consequently, the nonselective genomewide approach can be combined in a synergistic man-

ner with the more conventional detection strategies for selective targets to achieve better cancer management.

Given that plasma DNA in cancer patients consists of a mixture of tumor and nontumor DNA, with the former being present as a minority population in many patients, one needs to perform relatively deep sequencing to obtain the necessary analytical power. Thus, as currently implemented, the shotgun sequencing approach we have described is relatively expensive compared with many targeted approaches (8). With the continuing and, thus far, rapid reduction in sequencing costs, cost is not expected to be as much of an obstacle in the near future. One can also envision the possibility of using shotgun sequencing of a plasma sample upon presentation of a cancer patient, mining the shotgun sequencing data of plasma DNA for copy number aberrations and SNVs, and then specifically targeting such sequences for serial monitoring purposes. In this regard, solution-phase hybridization-based capture approaches have already been successfully used to enrich plasma DNA for noninvasive prenatal diagnosis (34, 35). Hence, we see that the targeted and shotgun approaches can be used in combination for cancer detection and monitoring.

In summary, we have demonstrated the performance of cancer genome scanning via shotgun sequencing of plasma DNA. This development has numerous clinical and research applications. It would therefore be a research priority to evaluate this approach on a large cohort of cancers of multiple types.

Author Contributions: All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design,

acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.

Authors' Disclosures or Potential Conflicts of Interest: Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:

Employment or Leadership: Y.M.D. Lo, *Clinical Chemistry*, AACC. **Consultant or Advisory Role:** R.W.K. Chiu, Sequenom; Y.M.D. Lo, Sequenom.

Stock Ownership: R.W.K. Chiu, Sequenom; Y.M.D. Lo, Sequenom. **Honoraria:** R.W.K. Chiu, Life Technologies (travel grants) and Illumina (travel grants); Y.M.D. Lo, Illumina and Life Technologies.

Research Funding: K.C.A. Chan, Hong Kong Research Grants Council Theme-based Research Scheme (T12-CUHK05/10); H. Sun, Hong Kong Research Grants Council Theme-based Research Scheme (T12-CUHK05/10); A.T.C. Chan, Hong Kong Research Grants Council Theme-based Research Scheme (T12-CUHK05/10), and the Innovation and Technology Fund under the State Key Laboratory Programme; R.W.K. Chiu, Hong Kong Research Grants Council Theme-based Research Scheme (T12-CUHK05/10) and S.K. Yee Foundation; Y.M.D. Lo, Hong Kong Research Grants Council Theme-based Research Scheme (T12-CUHK05/10), S.K. Yee Foundation, Innovation and Technology Fund under the State Key Laboratory Programme, and an endowed chair from the Li Ka Shing Foundation.

Expert Testimony: None declared.

Patents: K.C.A. Chan, P. Jiang, R.W.K. Chiu, and Y.M.D. Lo have all declared the same set of 5 US patent applications filed on this work: 13/308473, 61/662878, 61/682725, 61/695795, and 61/711172.

Data and Materials Availability: Sequence data and genotype data have been deposited at the European Genome-Phenome Archive (EGA, <http://www.ebi.ac.uk/ega/>), which is hosted by the European Bioinformatics Institute (EBI), under accession number EGAS00001000370.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or preparation or approval of manuscript.

Acknowledgments: We thank L. Chan, Y. Jin, C. Lee, K. Chow, S.-W. Yeung, X. Su, and C. Chan for technical assistance.

References

- Schwarzenbach H, Hoon DS, Pantel K. Cell-free nucleic acids as biomarkers in cancer patients. *Nat Rev Cancer* 2011;11:426–37.
- McDermott U, Downing JR, Stratton MR. Genomics and the continuum of cancer care. *N Engl J Med* 2011;364:340–50.
- Nawroz H, Koch W, Anker P, Stroun M, Sidransky D. Microsatellite alterations in serum DNA of head and neck cancer patients. *Nat Med* 1996;2:1035–7.
- Chen XQ, Stroun M, Magnenat JL, Nicod LP, Kurt AM, Lyautey J, et al. Microsatellite alterations in plasma DNA of small cell lung cancer patients. *Nat Med* 1996;2:1033–5.
- Yung TK, Chan KCA, Mok TS, Tong J, To KF, Lo YMD. Single-molecule detection of epidermal growth factor receptor mutations in plasma by microfluidics digital PCR in non-small cell lung cancer patients. *Clin Cancer Res* 2009;15:2076–84.
- Diehl F, Li M, Dressman D, He Y, Shen D, Szabo S, et al. Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc Natl Acad Sci U S A* 2005;102:16368–73.
- Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, et al. Circulating mutant DNA to assess tumor dynamics. *Nat Med* 2008;14:985–90.
- Forshe T, Murtaza M, Parkinson C, Gale D, Tsui DW, Kaper F, et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci Transl Med* 2012;4:136–68.
- Diaz LA Jr, Williams RT, Wu J, Kinde I, Hecht JR, Berlin J, et al. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* 2012;486:537–40.
- Wong IH, Lo YMD, Zhang J, Liew CT, Ng MH, Wong N, et al. Detection of aberrant p16 methylation in the plasma and serum of liver cancer patients. *Cancer Res* 1999;59:71–3.
- Begum S, Brait M, Dasgupta S, Ostrow KL, Zahurak M, Carvalho AL, et al. An epigenetic marker panel for detection of lung cancer using cell-free serum DNA. *Clin Cancer Res* 2011;17:4494–503.
- Lo YMD, Chan LYS, Lo KW, Leung SF, Zhang J, Chan AT, et al. Quantitative analysis of cell-free Epstein-Barr virus DNA in plasma of patients with nasopharyngeal carcinoma. *Cancer Res* 1999;59:1188–91.
- Leary RJ, Kinde I, Diehl F, Schmidt K, Clouser C, Duncan C, et al. Development of personalized tumor biomarker using massively parallel sequencing. *Sci Transl Med* 2010;2:20–14.
- McBride DJ, Orpana AK, Sotiriou C, Joensuu H, Stephens PJ, Mudie LJ, et al. Use of cancer-specific genomic rearrangements to quantify disease burden in plasma from patients with solid tumors. *Genes Chromosomes Cancer* 2010;49:1062–9.
- Lo YMD, Chiu RWK. Genomic analysis of fetal nucleic acids in maternal blood. *Annu Rev Genomics Hum Genet* 2012;13:285–306.
- Chiu RWK, Chan KCA, Gao Y, Lau VYM, Zheng W, Leung TY, et al. Noninvasive prenatal diagnosis

- of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. *Proc Natl Acad Sci U S A* 2008;105:20458–63.
17. Palomaki GE, Kloza EM, Lambert-Messerlian GM, Haddow JE, Neveux LM, Ehrich M, et al. DNA sequencing of maternal plasma to detect Down syndrome: an international clinical validation study. *Genet Med* 2011;13:913–20.
 18. Bianchi DW, Platt LD, Goldberg JD, Abuhamad AZ, Sehnert AJ, Rava RP. Genome-wide fetal aneuploidy detection by maternal plasma DNA sequencing. *Obstet Gynecol* 2012;119:890–901.
 19. Lo YMD, Chan KCA, Sun H, Chen EZ, Jiang P, Lun FMF, et al. Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci Transl Med* 2010;2:61–91.
 20. Kitzman JO, Snyder MW, Ventura M, Lewis AP, Qiu R, Simmons LE, et al. Noninvasive whole-genome sequencing of a human fetus. *Sci Transl Med* 2012;4:137–76.
 21. Chan KCA, Zhang J, Hui AB, Wong N, Lau TK, Leung TN, et al. Size distributions of maternal and fetal DNA in maternal plasma. *Clin Chem* 2004;50:88–92.
 22. Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 2009;25:1966–7.
 23. Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardisson D, Mannucci PM, et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet* 2009;41:334–41.
 24. Korn JM, Kuruvilla FG, McCarroll SA, Wysoker A, Nemesh J, Cawley S, et al. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet* 2008;40:1253–60.
 25. De Keersmaecker K, Real PJ, Gatta GD, Palomero T, Sulis ML, Tosello V, et al. The TLX1 oncogene drives aneuploidy in T cell transformation. *Nat Med* 2010;16:1321–7.
 26. Dohm JC, Lottaz C, Borodina T, Himmelbauer H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res* 2008;36:e105.
 27. Alkan C, Kidd JM, Marques-Bonet T, Aksay G, Antonacci F, Hormozdiari F, et al. Personalized copy number and segmental duplication maps using next-generation sequencing. *Nat Genet* 2009;41:1061–7.
 28. Chen EZ, Chiu RW, Sun H, Akolekar R, Chan KC, Leung TY, et al. Noninvasive prenatal diagnosis of fetal trisomy 18 and trisomy 13 by maternal plasma DNA sequencing. *PLoS One* 2011;6:e21791.
 29. Minoche AE, Dohm JC, Himmelbauer H. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biol* 2011;12:R112.
 30. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. *Genome Res* 2009;19:1639–45.
 31. Chiu RW, Akolekar R, Zheng YW, Leung TY, Sun H, Chan KC, et al. Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. *BMJ* 2011;342:c7401.
 32. Lo YMD, Lun FMF, Chan KCA, Tsui NBY, Chong KC, Lau TK, et al. Digital PCR for the molecular detection of fetal chromosomal aneuploidy. *Proc Natl Acad Sci U S A* 2007;104:13116–21.
 33. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* 2012;366:883–92.
 34. Liao GJ, Lun FM, Zheng YW, Chan KC, Leung TY, Lau TK, et al. Targeted massively parallel sequencing of maternal plasma DNA permits efficient and unbiased detection of fetal alleles. *Clin Chem* 2011;57:92–101.
 35. Lam KW, Jiang P, Liao GJ, Chan KCA, Leung TY, Chiu RW, Lo YMD. Noninvasive prenatal diagnosis of monogenic diseases by targeted massively parallel sequencing of maternal plasma: application to β -thalassemia. *Clin Chem* 2012;58:1467–75.