CANCER GENETICS

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.

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

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 × 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 (interchromosomal 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

Downloaded from https://www.science.org on November 05, 2022

Foresight Ex 1002
Foresight v Personalis
IPR2023-00546

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.

The second approach combined mate-paired tag data with copy number alterations identified by analyses of individual 25-bp tags. Tumor-specific copy number alterations are often associated with de novo rearrangements (*23*), and the boundaries of such alterations

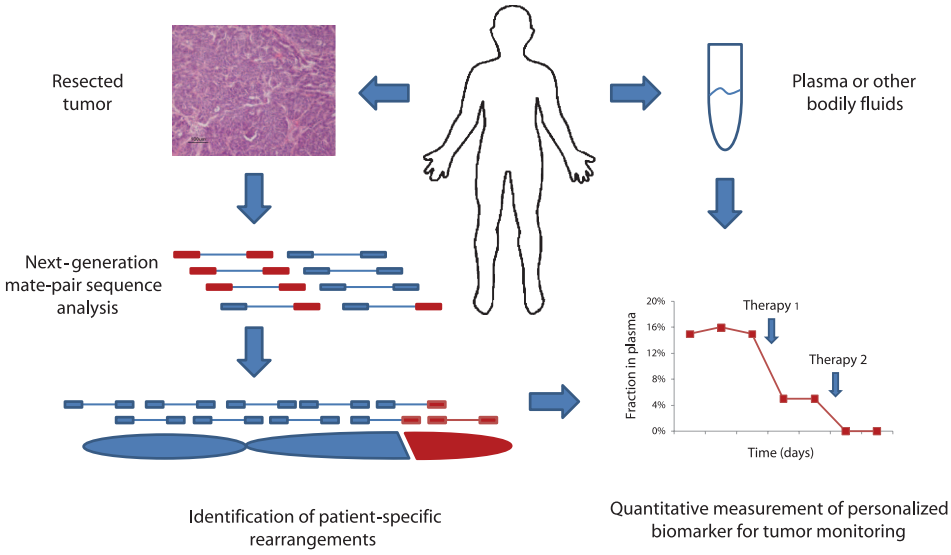


Fig. 1. Schematic of PARE approach. The method is based on next-generation mate-paired analysis of resected tumor DNA to identify individualized tumor-specific rearrangements. Such alterations are used to develop PCR-based quantitative analyses for personalized tumor monitoring of plasma samples or other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

Downloaded from https://www.science.org on November 05, 2022

Foresight Ex 1002
Foresight v Personalis
IPR2023-00546

Table 2. Summary of rearrangements identified in tumor samples.

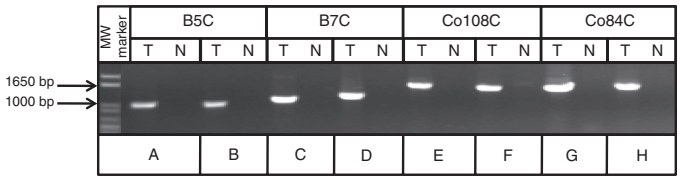
Sample	Rearrangement type		Total rearrangements	Tested rearrangements	Confirmed somatic rearrangements
	Intrachromosomal	Interchromosomal			
		Tumor and normal libraries			
B5	7	4	11	7	5 (71%)
B7	17	4	21	16	15 (94%)
Co84	0	7	7	6	4 (67%)
Co108	6	12	18	13	11 (85%)
		Tumor libraries			
Hx402	7	2	9	9	4 (44%)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

A



B

Alteration	Sample	Forward tag		Reverse tag		Rearrangement type
		Chromosome	Position	Chromosome	Position	
A	B5	11	57,714,122	8	-48,889,091	Interchromosomal
B	B5	20	29,591,563	18	-19,142,112	Interchromosomal
C	B7	5	38,284,430	10	-44,715,202	Interchromosomal
D	B7	6	-90,854,024	6	-106,401,711	Intrachromosomal
E	Co108	3	60,573,270	3	60,473,213	Intrachromosomal
F	Co108	4	81,934,151	15	54,039,041	Interchromosomal
G	Co84	8	128,442,121	19	49,144,200	Interchromosomal
H	Co84	13	-109,267,878	11	-34,791,081	Interchromosomal

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.

Downloaded from https://www.science.org on November 05, 2022

Foresight Ex 1002
Foresight v Personalis
IPR2023-00546

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 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.

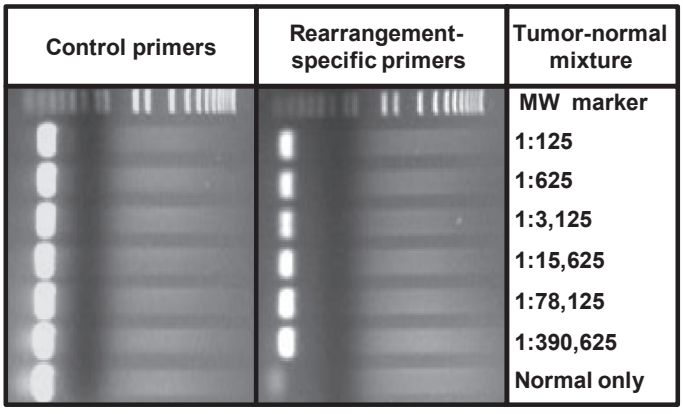


Fig. 3. Detection of tumor-specific rearrangements in mixtures of tumor and normal DNA. Decreasing amounts of tumor DNA were mixed with increasing amounts of normal tissue DNA (300 ng total) and were used as template molecules for PCR using chromosome 4:8 translocation-specific primers or chromosome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

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-

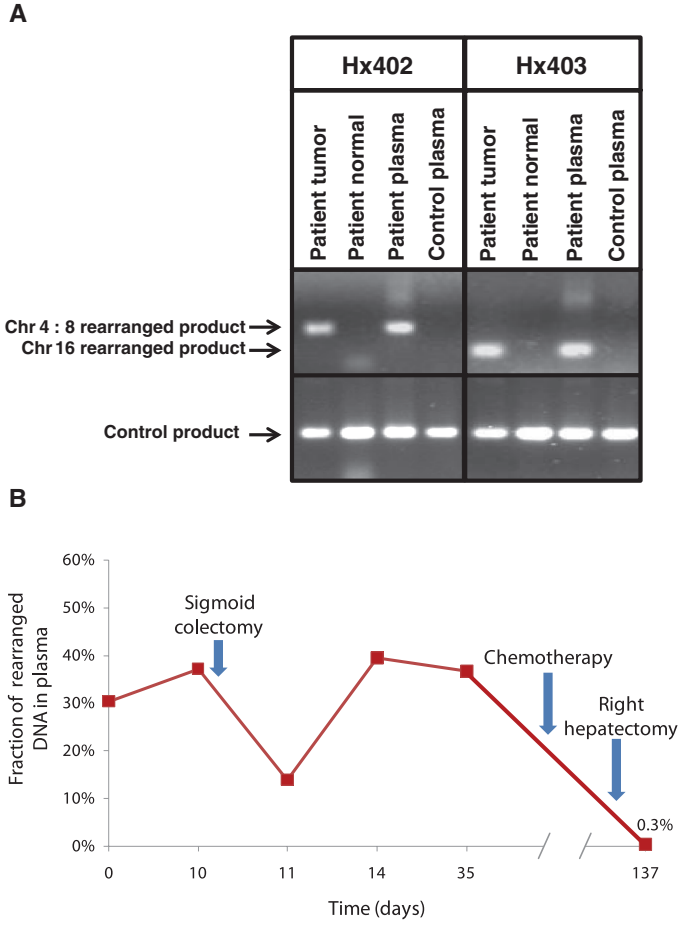


Fig. 4. Detection of tumor-specific rearrangements in plasma of cancer patients. (**A**) The identified chromosome 4:8 and 16 rearrangements were used to design PCR primers spanning breakpoints and to amplify rearranged DNA from tumor tissue and plasma from patients Hx402 and Hx403, respectively. A plasma 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 November 05, 2022

Foresight Ex 1002
Foresight v Personalis
IPR2023-00546

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 management 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 NlaIII and SacI 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

Downloaded from https://www.science.org on November 05, 2022

Foresight Ex 1002
Foresight v Personalis
IPR2023-00546

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 translocation 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. Segraves, 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,

Downloaded from https://www.science.org on November 05, 2022

Foresight Ex 1002
Foresight v Personalis
IPR2023-00546

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. Hurles, 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. Clee, M. A. Quail, A. Cox, C. Brown, R. Durbin, M. E. Hurles, 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. Gottimukkala, 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. Børresen-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).

Downloaded from https://www.science.org on November 05, 2022

Foresight Ex 1002
Foresight v Personalis
IPR2023-00546

Science Translational Medicine

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

Downloaded from https://www.science.org on November 05, 2022

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

Foresight Ex 1002
Foresight v Personalis
IPR2023-00546