REVIEW

Whole exome sequencing for cancer – is there evidence of clinical utility?

Alka Malhotra
Susan Levine
Diane Allingham-Hawkins

Genetic Test Evaluation Program,
Winifred S Hayes, Inc., Lansdale,
PA, USA

Video abstract

Point your SmartPhone at the code above. If you have a QR code reader the video abstract will appear. Or use:
http://dvpr.es/1r47PQV

Correspondence: Alka Malhotra; Diane
Allingham-Hawkins
Genetic Test Evaluation Program,
Winifred S Hayes, Inc., 157 South Broad
Street, Lansdale, PA 19446 USA
Tel +1 215 855 0615
Email amalhotra10@gmail.com;
dahawkins@hayesinc.com

Background: In recent years, whole exome sequencing (WES), which allows detection of 85% of disease-causing variants, has been used to compare tumor and normal DNA to allow the identification of variants specific to the tumor. Genetic changes in cancer are increasingly used for diagnosis and may guide treatment decisions. In this paper, we explore whether there is evidence that WES improves outcomes for patients with cancer.

Methods: Published evidence was evaluated using a methodology that combines the analytical validity, clinical validity, clinical utility and ethical, legal, and social implications (ACCE) model for genetic test evaluations with internationally accepted health technology assessment methodology. Conclusions were based on peer-reviewed published studies of >10 patients, with ≥3 studies for a given phenotype.

Results: WES has been conducted most extensively (seven studies to date) in breast cancer patients, with fewer studies of other types of cancers (eg, leukemia, prostate cancer, and ovarian cancer). Studies evaluating somatic alterations showed high intratumor and intertumor heterogeneity. In addition, both novel and previously implicated variants were identified. However, only three studies with >10 individuals have shown potential for clinical utility of WES; whereby, variants identified through WES may determine response to drug treatment.

Conclusion: Despite evidence for clinical validity of WES in cancers, clinical utility is very limited and needs to be further evaluated in large clinical studies.

Keywords: next-generation sequencing, exons, evidence-based, ACCE model, health technology assessment

Introduction

Cancers are caused by the accumulation of genetic alterations that may lead to the dysfunction of regulation of cell growth, resulting in the development of tumors.[1] According to the American Cancer Society, it is estimated that >1.6 million individuals were diagnosed with cancer in the US in 2013 alone, with >580,000 resulting deaths. In the US, the most common new cases of cancers included breast cancer, prostate cancer, lung cancer, and colon cancer.[2]

In the past decade, there have been significant developments in next-generation technologies to sequence DNA rapidly.[3,4] Massively parallel sequencing not only generates data from the entire genome in a short period of time, but it also contributes to cost reduction. Whole exome sequencing (WES) involves sequencing of all coding regions (exons) in the genome. Compared with whole genome sequencing (WGS), WES facilitates handling of data and also generates higher-quality data, since it can be performed at greater sequencing depth. The optimal sequencing depth depends on a

number of factors, including, but not limited to: the region(s) being sequenced; the sample source; and the algorithm used to assemble and analyze the results. While exons represent just 1% of the genome, they account for approximately 85% of disease-causing variants, specifically for Mendelian traits;[5] however, the percentage for complex traits is not known.[6]

In recent years, WES has been used to compare tumor DNA and normal DNA to identify variants specific to the tumor. WES of tumor DNA requires greater sequence depth than normal DNA to identify variants that are present in tumor cells only. These analyses may provide information about genes with driver variants (those with an effect on cancer development) versus passenger variants (those without an effect on cancer development).[7] Through identification of potentially deleterious variants, WES may provide information about potential new avenues for diagnosis and treatment.

While clinical exome sequencing is being offered by a number of laboratories, WES specifically for cancer is currently offered by two Clinical Laboratory Improvement Amendments-certified laboratories in the US. The Baylor College of Medicine Medical Genetics Laboratories offers the Cancer Exome Sequencing test,[8] and Personal Genome Diagnostics Inc., (Baltimore, MD, USA) offers the Cancer *Complete*™ test.[9] Using proprietary methods, Personal Genome Diagnostics Inc. also conducts quality assessment and evaluation of genes and pathways of interest, identified through WES.[9]

This systematic review evaluates the available published evidence about the use of WES for cancer indications, with emphasis on analytical validity, clinical validity, and clinical utility.

Methods
Search strategy
Evidence evaluated for this report was obtained primarily from a search of the peer-reviewed literature in PubMed and Embase performed on June 3, 2013. Search terms included "exome" AND "sequencing" AND "cancer" AND ("diagnosis" OR "clinical" OR "utility" OR "validity"). Limits used were English language, human, and published since January 1, 1996. Additional relevant citations were also selected from the bibliographies of retrieved references.

Evidence evaluation
The evidence analysis used for this review is based on the ACCE model (http://www.cdc.gov/genomics/gtesting/ACCE/index.htm) that was developed by the Centers for Disease

Control and Prevention.[10,11] The ACCE model takes its name from an abbreviation of its main components: analytical validity; clinical validity; clinical utility; and ethical, legal, and social implications. The model is widely used to allow the performance of rapid evaluations of genetic tests.[12]

For this paper, the analytical validity is defined as the ability of WES technology to measure accurately and reliably the sequences of interest. Clinical validity is defined as the ability of WES to detect or predict the associated disorder or phenotype, while clinical utility focuses on what needs to be considered when evaluating the risks and benefits of introducing a genetic test into routine practice, and it is based on studies designed to investigate whether there are improvements in health outcomes as a result of using the genetic test in clinical practice. In this review, we focus on these components of the ACCE model, with less emphasis on the ethical, legal, and social implications.

Results/evidence overview
Analytical validity
Commercial laboratories offering WES mainly use the Agilent SureSelect All Exome (Agilent Technologies, Santa Clara, CA, USA) or NimbleGen Sequence Capture (SeqCap) Human Exome (Hoffmann-La Roche Ltd, Basel, Switzerland) platforms for exome capture. The analytical validity of these methods has been assessed by four studies comparing various platforms to detect variants through WES. Three studies utilized data from healthy individuals and one study evaluated data from cancer patients; however, none of the information provided was specific to cancer. In the analyses, various parameters were evaluated, including the following:

- Enrichment efficiency – proportion of base pair reads falling in the target region to the total base pair reads in any region of the genome.
- Genotype sensitivity – probability to accurately determine the genotypes in the target regions.
- Genotype concordance – accuracy of overlap of genotype calls between the current analysis and previously identified genotype information.
- Coverage – overlap with existing publicly available databases with protein-coding variant information, as well as detection percentage depending on the depth of sequencing.

Table 1 presents comparisons between the different exome enrichment platforms, some that are currently being used by commercial laboratories in the US. Two studies were conducted in individuals of Asian descent,[13,14] and two

Table 1 Comparison of exome enrichment platforms

Reference	Sample	Platforms tested	Methods	Results
Asan et al (2011)[13]	One healthy Asian man who had previous WGS data available	Agilent SureSelect Human All Exon 50 Mb[a] NimbleGen SeqCap Ez[b] NimbleGen SeqCap Human Exome 2.1 M Array[b]	DNA was extracted from peripheral blood. Exomes were captured in two replicates for each platform. Sequencing was performed using Illumina HiSeq 2000 platform. Exome capture by the three platforms was compared with publicly available variant information in protein-coding genes from three databases: CCDS, refGen, and EnsemblGen. Genotype data were compared with WGS data and Illumina 1 M Infinium GWAS chip.	In most cases, results of SeqCap EZ and 2.1 M Array were similar; therefore, only one set of results is presented. SureSelect: Genotype sensitivity: 64%–85% at 20×–50× depth Enrichment efficiency: average across replicates =56.4% at >30× depth Coverage: 80.6% of publicly available data; 59% coverage at 10×–50× depth Genotype concordance: >99% NimbleGen: Genotype sensitivity: 72%–91% at 20×–50× depth Enrichment efficiency:average across replicates =54.8% at >30× depth Coverage: 75.9% of publicly available data; 70%–79% coverage at 10×–50× depth Genotype concordance: >99%
Clark et al (2011)[15]	One healthy European individual	Agilent SureSelect Human All Exon 50 Mb[c] NimbleGen SeqCap EZ Exome Library v2.0[b] Illumina TruSeq Exome Enrichment[c]	DNA was extracted from peripheral blood mononuclear cells. Sequencing was performed using Illumina HiSeq 2000 platform. Genotype data were compared with Illumina 1 M Duo SNP chip.	SureSelect: Enrichment efficiency: 89.6% at ≥10× depth Coverage: >80% at 25× depth Genotype concordance: 99.3% SeqCap EZ: Enrichment efficiency: 96.8% at ≥10× depth Coverage: >80% at 25× depth Genotype concordance: 99.5% TrueSeq: Enrichment efficiency: 90.0% at ≥10× depth Coverage: >80% at 25× depth Genotype concordance: 99.2%
Hedges et al (2011)[16]	18 healthy American individuals (eleven women; seven men)	Agilent SureSelect[a] NimbleGen SeqCap[b] RainDance™ Technology (multiplex PCR-based)[d]	DNA was extracted from peripheral blood leukocytes in 16 individuals and two cell lines. Variants in a common ~0.8 Mb were evaluated. Exons, 5 kb flanking regions, and additional evolutionarily conserved regions were assessed. 18 individuals were divided among three platforms; in addition, six randomly selected individuals were tested across platforms. Sequencing was performed using ABI SOLID3+ platform. Genotype data were compared with Illumina 1 M Infinium GWAS chip.	SureSelect: Enrichment efficiency: average =60.8% Coverage: >90% at 10× depth Genotype concordance: average =99.8% at >20× depth SeqCap EZ: Enrichment efficiency: average =53.3% Coverage: >90% at 10× depth Genotype concordance: average =99.7% at >20× depth RainDance: Enrichment efficiency: average =52.5% Coverage: >90% at 10× depth Genotype concordance: average =98.3% at >20× depth

(Continued)

Table I (*Continued*)

Reference	Sample	Platforms tested	Methods	Results
Guo et al (2012)[14]	28 Chinese individuals with breast cancer – very early onset (age 22–32 years) or early onset (38–41 years) and a first-degree relative with breast cancer	Agilent SureSelect v1[a] (n=22) Illumina TruSeq Exome Enrichment[c] (n=8)	Sequencing was performed using Illumina HiSeq 2000 platform (Illumina samples) or GA II (Agilent samples). Exome capture by two platforms was compared with CCDS database. Comparisons were conducted within target region and outside target region (since flanking regions are also usually sequenced); however, only results of within-target region are presented.	SureSelect: Coverage: 97% of sites were covered with at least one read; average number of reads was 68.9 million per sample at 45× median depth; 98.3% of CCDS variants were detected Consistency rate (similar to genotype concordance): >99% at 20× depth TrueSeq: Coverage: ~97% of sites were covered with at least one read; average number of reads was 93.8 million per sample at 48× median depth; 96.5% of CCDS variants were detected Consistency rate: >99% at 20× depth

Notes: Adapted with permission from Winifred S Hayes, Inc. Manufacturer details are as follows: [a]Agilent Technologies, Santa Clara, CA, USA; [b]F Hoffmann-La Roche Ltd., Basel, Switzerland; [c]Illumina Inc., San Diego, CA, USA; [d]RainDance Technologies, Inc., Billerica, MA, USA.
Abbreviations: CCDS, Consensus Coding Sequence; GA II, Genome Analyzer II; GWAS, genome-wide association study; PCR, polymerase chain reaction; SeqCap, Sequence Capture; SNP, single nucleotide polymorphism; WGS, whole genome sequencing.

studies were of white individuals.[15,16] For all studies, a high genotype concordance (>98%) was estimated.[13–16] However, overall coverage rates varied between studies and platforms, ranging from >80% in two studies[15,16] to 59% (for Agilent SureSelect [Agilent Technologies]) in one study.[13] In addition, where available, enrichment efficiency also varied among studies, ranging from 56.4%–89.6% for the Agilent Sure-Select (Agilent Technologies) platform.[13,15,16] This variation may be explained by the different depths of coverage used by each study.

Clinical validity

The use of WES has been reported for a number of different cancers, with a majority of studies being either case reports or having sample sizes < ten individuals, which is too small to draw meaningful conclusions. This section describes clinical validity data obtained from at least three studies of the same or similar cancer type, with each of the studies including > ten individuals.

Breast cancer

Seven studies with > ten individuals evaluated the clinical validity of WES in breast cancer (summarized in Table 2). The overall proportion of missense variants in unrelated individuals was similar in different studies, ranging from 61.2%–65.4%.[17–21] In addition, studies identified varying proportions of variants and candidate genes depending on the subtype; for example, the *tumor protein p53* (*TP53*) gene had a high proportion of variants in luminal B individuals (P=0.04), while the *mitogen-activated protein kinase kinase kinase 1* (*MAP3K1*) gene had a high proportion of variants in luminal A individuals (P=0.02).[19] In all studies, both previously identified genes – as well as those not previously detected in breast cancer studies – were identified with a larger number of variants than expected by chance.[17,19–23]

Two studies were specifically conducted in families that were negative for the *breast cancer 1, early-onset* (*BRCA1*) gene and the *breast cancer 2, early-onset* (*BRCA2*) gene.[22,23] One study identified the *Bloom syndrome, RecQ helicase-like* (*BLM*) and *Fanconi anemia, complementation group C* (*FANCC*) genes with higher frequency of variants that may have an effect on breast cancer development;[23] whereas, the second study identified eleven candidate genes.[22]

Leukemia and lymphoma

Table 3 presents results of WES in two types of leukemia – two studies for chronic lymphocytic leukemia (CLL)[24,25] and one study for acute myeloid leukemia (AML).[26] While all

Table 2 WES in breast cancer patients

Reference	Population	Methodology	Results	Gene information/follow-up analysis
Banerji et al (2012)[17]	108 Mexican (n=56) and Vietnamese (n=52) patients with diverse subtypes: 38 (35.2%) luminal A; 22 (20.4%) luminal B; 21 (19.4%) HER2-enriched; 13 (12%) basal/basal-like; 9 (8.3%) normal-like; and 5 (4.6%) unknown. Of these, data from 103 and 5 patients were used for WES and WGS, respectively.	Frozen breast cancer samples and normal matched controls (peripheral blood from the Mexican patients and normal breast tissue from Vietnamese patients) were used for WES. Exome was captured using a solution hybrid selection method. 189,980 exons across 33 Mb were sequenced using Illumina GA II platform.[a] Validation of variants was performed using three different sequencing technologies.	4,985 somatic variants were detected, with an overall rate of 1.66 variants per Mb. Variant type: Missense: 3,153 (63.2%) Silent: 1,157 (23.2%) Nonsense: 242 (4.9%) Splice site: 97 (1.9%) Deletions: 194 (3.9%) Insertions: 110 (2.2%) Nonsilent variant rate was 1.27 per Mb. Of 494 variants tested for validation, 94% were confirmed.	Follow-up analysis of variant frequencies identified six genes showing variants recurring in tumor samples when compared with normal samples ($P \leq 0.00001$).
Cancer Genome Atlas Research Network (2012)[18]	825 patients from US with no prior treatment with chemotherapy or radiotherapy. Of these, information on 510 tumors from 507 patients was used in this study:* 225 (44.4%) luminal A; 126 (24.9%) luminal B; 57 (11.2%) HER2-enriched; 93 (18.3%) basal-like; 6 (1.2%) unknown.	Frozen tumors and matched normal controls (blood sample or normal breast tissue) were used. Hybridization was performed using Agilent SureSelect All Exome kit v2.0[b] or NimbleGen SeqCap Human Exome v2.0.[c] Illumina HiSeq 2000[a] was used for sequencing.	30,626 somatic variants were detected. Variant type: Missense: 19,045 (62.2%) Silent: 6,486 (21.2%) Nonsense: 1,437 (4.7%) Splice site: 506 (1.7%) Indels: 2,302 (7.5%) Stop codon read-through: 26 (0.1%) Of the missense variants, 9,484 (49.8%) had a high probability of being deleterious. In addition, average nonsilent variant rate was ~1.49 per Mb.	Follow-up analysis of variant frequencies identified 35 genes showing a higher number of variants than expected by chance ($P < 0.05$ for all tumor types).
Ellis et al (2012)[19]	77 American ER+ patients undergoing aromatase inhibitor treatment. Of these, data from 46 and 31 patients were used for WGS and WES, respectively.	Fresh frozen tumor and matched normal peripheral blood samples were used. Hybridization was performed using Roche NimbleGen Sequence Capture EZ Human Exome Library v2.0 kit.[c] An Illumina platform[a] (specific technology not specified) was used for sequencing.	3,355 coding somatic variants were detected using both WES and WGS. Of these, 1,371 were identified through WES alone. Variant type[#] (of the total 3,355): Missense: 2,145 (63.9%) Nonsense: 178 (5.3%) Splice site: 69 (2.1%) Indels: 146 (4.4%) Stop codon read-through: 6 (0.2%) Of missense variants, 1,551 (72.3%) had a high probability of being deleterious. Variant rate was 1.05 per Mb.	Follow-up analysis of variant frequencies identified 18 genes showing higher number of variants than expected by chance ($P \leq 1.79 \times 10^{-4}$).

(*Continued*)

Table 2 (*Continued*)

Reference	Population	Methodology	Results	Gene information/follow-up analysis
Shah et al (2012)[20]	149 triple-negative (ie, ER−, PR−, and not HER2-enriched) patients from the UK and Canada. Of the 149 patients, WES was performed for 58 patients.[§]	Fresh frozen tumors and matched normal samples (breast tissue or peripheral blood lymphocytes) were used. Exome enrichment was performed using Agilent SureSelect Human All Exome v1 kit.[b] Illumina GA II platform[a] was used for sequencing. Variants were validated by targeted deep sequencing.	2,414 detected SNV somatic variants were validated (information on overall variant frequencies was not available). Variant type (of total 2,414): Missense: 2,212 (91.6%) Splice site: 43 (1.8%) Stop (truncating) variant: 159 (6.6%)	Follow-up analysis of variant frequencies identified 17 genes showing a higher number of variants than expected by chance (P≤0.005). In addition, P<0.05 for 238 genes.
Stephens et al (2012)[21]	100 patients from Europe, Asia, and Australia: 79 ER+ and 21 ER−.	Tumors and normal tissue from the same individual were used. Exome enrichment was performed using Agilent SureSelect Human All Exome 50 Mb kit.[b] Illumina HiSeq or GA II platform[a] was used for sequencing.	7,241 somatic variants were detected. Variant type: Missense: 4,737 (65.4%) Silent: 1,637 (22.6%) Nonsense: 422 (5.8%) Splice site: 158 (2.2%) Deletions: 206 (2.8%) Insertions: 71 (0.1%) Stop codon read-through: 8 (0.01%) Variant rates per Mb were not provided.	>40 previously identified and novel candidate genes showed presence of protein-altering somatic variants.
Thompson et al (2012)[23]	33 patients from 15 families from Australia and New Zealand in whom at least one family member had BRCA1 and BRCA2 variant analysis, but results were negative.	Blood from patients and age- and ethnicity-matched controls (from the UK) was obtained. Exome enrichment was performed using Agilent SureSelect All Exome kit v2.0 (or 50 Mb)[b] or NimbleGen SeqCap 2.1 M.[c] Illumina HiSeq or GA II platform[a] was used for sequencing.	536 deleterious variants were identified. Variant type: Frameshift: 331 (61.8%) Splice site: 79 (14.7%) Nonsense: 103 (19.2%) Indel: 23 (4.3%) An average of 35 novel deleterious variants was detected per individual, with an average of nine variants shared by all family members. In addition, 284 nonsynonymous variants were detected per individual.	Follow-up analysis of variant frequencies identified two genes with potentially deleterious variants. Sanger sequencing validated these results.
Gracia-Aznarez et al (2013)[22]	Families from five countries participated in this study. Eleven patients from seven families negative for BRCA1 and BRCA2 variants with at least six breast cancer cases in the family were selected. In addition, seven control cell lines were used.	Exome enrichment was performed using Agilent SureSelect Human All Exome 50 Mb kit.[b] The Illumina HiSeq or GA II platform[a] was used for sequencing. Variants were validated through comparison with previously published data, and by using Sanger sequencing.	28,249 SNPs and 31,526 indels were detected. Following stringent filtering criteria (which included absence in controls, presence in at least two family members, and potential gene function), a total of 67 and 14 candidate SNPs and indels were identified, respectively. Following filtering, an average of 10 SNPs and two indels were detected per family, respectively.	Follow-up analysis of candidate variants identified eleven potential genes with protein-altering effects.

Notes: Adapted with permission from Winifred S Hayes, Inc. *Of the 825 individuals, WES was performed for 507 individuals. Data from the remaining 318 individuals were analyzed using other methods, including DNA methylation, microarray expression, and SNP analysis; [#]results are a combination of WES and WGS results unless specifically stated otherwise; [§]of the 149 individuals, WES was performed for 58 individuals. Data from the remaining individuals were analyzed using other methods, including RNA sequencing and SNP analysis. Manufacturer details are as follows: [a]Illumina Inc, San Diego, CA, USA; [b]Agilent Technologies, Santa Clara, CA, USA; [c]F Hoffmann–La Roche Ltd, Basel, Switzerland.
Abbreviations: BRCA1/2, breast cancer gene 1/2; ER+/−, estrogen receptor positive/negative; GA II, Genome Analyzer II; HER2, human epidermal growth factor receptor 2; indel, insertion/deletion; PR−, progesterone receptor negative; SeqCap, Sequence Capture; SNP, single nucleotide polymorphism; SNV, single nucleotide variation; WES, whole exome sequencing; WGS, whole genome sequencing.

Table 3 WES in leukemia patients

Reference	Cancer type and population	Methodology	Results	Gene information/follow-up analysis
Wang et al (2011)[25]	91 patients with CLL from US. Of the 91 patients, data from 88 were analyzed using WES, and three were analyzed using WGS.	Lymphocytes from blood or bone marrow were used for tumor samples. For normal tissue, epithelial cells, fibroblasts, or granulocytes were used. Exomes were captured using a solution hybrid selection method. The Illumina GA II platform[a] was used for sequencing.	Variant type (in protein-coding regions): Nonsynonymous: 1,838 Synonymous: 539 Missense: 1,503 (63.2% of total) Nonsense: 85 (3.6%) Splice site: 101 (4.2%) Insertions: 45 (1.9%) Deletions: 92 (3.9%) Stop codon read-through: 5 (0.2%) A somatic variant rate was estimated at 0.72±0.36 per Mb, with 20 nonsynonymous variants per patient.	Follow-up analysis of variant frequencies identified nine driver genes showing variants with potentially tumorigenesis effects. In addition, all variants in these genes were nonsynonymous ($P<5\times10^{-6}$).
Quesada et al (2012)[24]	105 patients from Spain with CLL: 45 with no variants in the IGHV region and 60 with variants in this region.	Fresh or frozen tumor samples and whole blood normal samples were used. Exome enrichment performed using Agilent SureSelect Human All Exon 50 Mb kit.[b] Illumina GA IIx platform[a] was used for sequencing.	1,246 somatic protein-altering variants were detected. Variant type: Nonsynonymous: 1,125 (90.3%) Splice site: 86 (6.9%) Frameshift: 35 (2.8%) Median of 45 variants per tumor detected, giving a variant rate of 0.9 variants per Mb. Larger number of protein-affecting variants were present in patients with variants in the IGHV region compared to those with no variants (12.8±0.7 versus 10.6±0.7; $P=0.038$).	78 genes were identified with potential protein-altering variants (with 60 genes having significantly higher number of variants than expected by chance; $P<0.05$). Of these, SF3B1 was further analyzed in 279 additional CLL patients, and was associated with shorter time to disease progression ($P=0.0001$) and lower 10-year survival rate ($P=0.002$).
Cancer Genome Atlas Research Network (2013)[26]	200 patients from US age 18–88 years with untreated de novo AML. WES was performed in 150 cases and WGS in 50 cases.	Exome capture was performed using a Roche NimbleGen kit.[c] Illumina HiSeq 2000 or GA IIx[a] was used for sequencing.	Total of 2,583 variants* were detected in coding sequences.# Variant type: Missense: 1,539 (59.6%) Nonsense: 117 (4.5%) Silent: 510 (19.7%) Splice site: 55 (2.1%) Deletions: 78 (3.0%) Insertions: 173 (6.7%)	Follow-up analysis of variant frequencies identified 23 genes with a variant rate higher than expected by chance (false discovery rate <0.05). In addition, 237 genes had variants in ≥2 individuals.

Notes: Adapted with permission from Winifred S Hayes, Inc. *The results are a combination of WES and WGS results, unless specifically stated otherwise; #there were slight discrepancies between the numbers presented in the text and Table S6 of the study. The table results are presented here. Manufacturer details are as follows: [a]Illumina Inc., San Diego, CA, USA; [b]Agilent Technologies, Santa Clara, CA, USA; [c]F Hoffmann-La Roche Ltd, Basel, Switzerland.

Abbreviations: AML, acute myeloid leukemia; CLL, chronic lymphocytic leukemia; GA II, Genome Analyzer II; IGHV, immunoglobulin heavy chain variable; SF3B1, splicing factor 3B, subunit 1 gene; WES, whole exome sequencing; WGS, whole genome sequencing.

three studies identified a large proportion of genes with variants with potential deleterious effects, these genes did not overlap between AML and CLL.[24–26] In the CLL studies, the *splicing factor 3b, subunit 1* (*SF3B1*) gene was identified as having a higher rate of protein-altering variants. Specifically, CLL patients with variants in this gene showed a shorter time to disease progression (P=0.0001) and lower 10-year survival rate (P=0.002).[24]

In addition to the previously mentioned leukemia studies, two WES studies have also been conducted in a related cancer, diffuse large B-cell lymphoma (DLBCL), which affects lymph nodes.[27,28]

- A DLBCL WES study was conducted by Lohr et al[27] (2012) using 55 frozen tumor samples and matched normal DNA from 55 US patients. Exomes were captured using a solution hybrid selection method followed by sequencing using the Illumina HiSeq platform (Illumina Inc., San Diego, CA, USA). Of the 55 tumors, six had very low variant rates and were, therefore, excluded from further analysis. In the remaining 49 samples, 6,233 total somatic variants were detected (missense, 4,093 [65.6%]; nonsense, 326 [5.2%]; silent, 1,549 [24.9%]; splice site, 123 [2.0%]; deletion, 98 [1.6%]; and insertion, 39 [0.6%]). A mean nonsynonymous variant rate was estimated at 3.2 per Mb (range, 0.6–7.8). Based on the WES results, 58 genes were identified as having a higher number of variants than expected (defined as a false discovery rate Q value <0.1; whereby, the analyses were corrected for multiple comparisons. In this study, the Q value indicated that <10% of significant tests were expected to result in a false-positive).[27]

- In a second study, as part of the Hematologic Malignancies Research Consortium, Zhang et al[28] (2013) analyzed a total of 73 frozen biopsies from US patients with DLBCL. These were divided into a discovery set (34 tumors and 34 matched unaffected bone marrow samples) and a validation set (the remaining 39 tumors). In addition, exomes from 21 DLBCL cell lines were sequenced. Exome enrichment was performed using the Agilent SureSelect All Exome kit (Agilent Technologies), and the Illumina HiSeq platform (Illumina Inc.) was used for sequencing. Following a first step of filtering, a total of 5,884 somatic variants were detected in the discovery set. Somatic variant frequencies were: 53.8%, missense; 1.1%, nonsense; 2.4%, frameshift; and 42.7%, silent. Upon further filtering, which included the removal of silent variants, 2,589 (discovery set) and 4,928 variants (discovery and validation sets combined) were identified.

Validation of variant detection was performed using three methods, resulting in 99% concordance in all cases. The follow-up of the WES results identified 322 strong candidate genes with recurrent variants in the DLBCL patients.[28]

Endometrial and ovarian cancer

Table 4 presents findings of four WES studies in endometrial and ovarian cancers, with the largest study performed by the Cancer Genome Atlas Research Network using 316 tumor samples. Three of the four studies showed the presence of different types of variants, including the missense, nonsense, splice site, deletions, and insertions.[29–31]

The number of variants per tumor varied greatly (some tumors with <100 variants and others having >3,000 variants), showing the high degree of genetic heterogeneity both within and among tumors.[31,32] In addition, the WES analyses resulted in the identification of nine to 15 genes per study, with more variants (including potentially deleterious variants in some studies) than would be expected by chance.[29–32]

Lung-related cancers

WES studies in lung cancers have been conducted for non-small-cell lung carcinoma (two studies)[33,34] and lung adenocarcinoma (one study).[35] These studies are described in Table 5. All the studies presented in this table used the Agilent SureSelect platform for exome enrichment, followed by the use of Illumina platforms (either Genome Analyzer II or HiSeq [Illumina Inc.]) for sequencing. The two non-small cell-lung carcinoma studies showed differences in the variant type, based on the subtype of the cancer;[33,34] for example, the ratio of nonsynonymous to synonymous variants was nominally statistically significantly higher in adenocarcinoma patients compared with squamous cell carcinoma patients (3.6 versus 2.8; P=0.05).[33] Follow-up of potential candidate genes identified >15 genes in all studies,[33–35] with some genes having a highly significant proportion of variants (P range; 2×10^{-14}–7×10^{-4}).[33]

WES in other types of cancer

WES studies have been conducted in other types of cancers, including colon and colorectal cancer,[36,37] prostate cancer,[38,39] and head and neck squamous cell carcinoma.[40,41] Since the methodology and criteria used to evaluate the results are similar to those described previously (ie, include assessment of the genotype landscape and, in some cases, identification of potential candidate genes to be further evaluated), these studies are not described in this review.

Table 4 WES in endometrial or ovarian cancer patients

Reference	Cancer type and population	Methodology	Results	Gene information/follow-up analysis
Cancer Genome Atlas Research Network (2011)[29]	489 patients from US with ovarian carcinoma. Of these, data from 316 patients were used.*	Frozen tumors and matched normal controls (lymphocytes, normal tissue, or previously extracted germline DNA from blood) were used. Exome enrichment was performed using Agilent SureSelect All Exome v2.0 kit[a] or NimbleGen SeqCap Human Exome v2.0.[b] Illumina GA IIx[c] (236 samples) and Applied Biosystems SOLiD™ platform[d] (80 samples) were used for sequencing.	19,356 somatic variants were detected. Variant type: Missense: 12,935 (66.8%) Nonsense: 817 (4.2%) Silent: 4,046 (20.9%) Splice site: 453 (2.3%) Deletions: 579 (3.0%) Insertions: 202 (1.0%) ~61 somatic variants per tumor were identified.	Follow-up studies based on results obtained from WES identified 15 genes with increased number of nonsynonymous or splice site variants than expected by chance.
Liang et al (2012)[30]	14 patients from US with endometrial cancer. Of these, data from 13 patients of different ethnicities were analyzed:[#] eight whites; one black; four Hispanics.	Frozen tumor samples and matched normal DNA from peripheral blood leukocytes were used. Exome enrichment was performed using Agilent SureSelect Human All Exon kit,[a] which captured 170,843 exons. Applied Biosystems SOLiD V3.0 Genome Analyzer[d] was used for sequencing.	1,164 variants were identified. Of these, 625 were nonsilent. Variant type (of 1,164 variants): Nonsynonymous: 576 (49.5%) Nonsense: 24 (2.1%) Small deletions: 127 (10.9%) Small insertions: 49 (4.2%) Mean of 89.5 somatic point variants per tumor were identified, giving a rate of 3.7/Mb.	Follow-up studies based on results obtained from WES identified 12 potential driver genes.
Kinde et al (2013)[31]	46 patients from Brazil and US: 22 with ovarian cancer and 24 with endometrial cancer. Of these, data from 22 endometrial tumors were used.[§]	Fresh frozen tumors and matched normal whole blood or adjacent tissue were used. Enrichment was performed using Agilent SureSelect Human Exon Kit v4.0.[a] Illumina GA IIx platform[c] was used for sequencing.	12,794 somatic variants were detected in the WES analysis. Variant type: nonsynonymous: 11,273 (88.1%) Splice site: 228 (1.8%) Deletions: 1,086 (8.5%) Insertions: 207 (1.6%) Ten tumors had <100 variants per tumor (range, 7–50) and 12 tumors had >100 variants (range, 164–4,629).	Follow-up studies based on results obtained from WES identified 15 genes with an increased number of variants than expected by chance.
Zhao et al (2013)[32]	57 patients from US with uterine serous carcinoma (endometrial cancer). Of these, 15 cell lines and 42 fresh frozen tumors were used. In addition, data from 34 controls were used.	Cell lines and fresh frozen tumors and matched normal tissue from same individual were used. Exome enrichment was performed using NimbleGen 2.1 M human exome array.[b] Illumina HiSeq platform[d] was used for sequencing. Only SNVs were assessed in this study.	Details of overall variant numbers and types are not available. Protein-altering variant rates varied greatly: <100 variants in 52 tumors and >3,000 variants in five tumors.	Follow-up studies based on results obtained from WES identified 14 genes with an increased number of potentially deleterious variants (P range, 1.01×10⁻⁴⁵–1.93×10⁻⁶) Of these, six genes had recurrent somatic variants.

Notes: Adapted with permission from Winifred S Hayes, Inc. *Of the 489 individuals, WES was performed for 316 individuals; data from the remaining 173 individuals were analyzed using other methods, including DNA methylation, microarray expression, and SNP analysis; [#]one tumor had an abnormally high variant frequency and was not included in the analyses; [§]in this study, data from only 22 individuals with endometrial cancer were analyzed. Results presented are for these individuals only. Manufacturer details are as follows: [a]Agilent Technologies, Santa Clara, CA, USA; [b]F Hoffmann-La Roche Ltd, Basel, Switzerland; [c]Illumina Inc, San Diego, CA, USA; [d]Thermo Fisher Scientific, Waltham, MA, USA.
Abbreviations: GA II, Genome Analyzer II; SeqCap, sequence capture; SNV, single nucleotide variant; WES, whole exome sequencing.

Table 5 WES in lung-related cancer patients

Reference	Cancer type and population	Methodology	Results	Gene information/follow-up analysis
Imielinski et al (2012)[35]	183 lung adenocarcinoma patients from the US and Canada: 118 heavy smokers; 17 light smokers; 17 never-smokers; 27 never-smokers; and 21 with unknown status.	Frozen tumors and normal adjacent tissue were used. WES was performed in 172 samples, and WGS was performed in 24 samples (23 samples overlapped in WES and WGS analyses). Enrichment was performed using Agilent SureSelect Human All Exon 50 Mb.[a] Illumina HiSeq platform[b] was used for sequencing.	77,736 somatic variants were detected. Variant type: Missense: 43,813 (56.4%) Nonsense: 3,504 (4.5%) Silent: 14,801 (19.0%) Splice site: 1,460 (1.9%) Deletions: 2,310 (3.0%) Insertions: 839 (1.1%) Mean variant rate was 11.9/Mb (range, 0.04–117.4). In addition, mean variant rates were 12.9/Mb and 2.9/Mb for smokers and never-smokers, respectively ($P=3.0\times10^{-9}$).	Follow-up studies based on results obtained from WES identified 25 genes with increased number of variants than expected by chance.
Liu et al (2012)[33]	31 NSCLC patients from US: 16 ADC; 12 SCC; one adenosquamous carcinoma; one BAC; and one ADC with BAC.	Tumors and matched normal tissues were analyzed. Enrichment was performed using Agilent SureSelect[a] to capture data from 212,919 exons. Illumina GA II and HiSeq 2000[b] platforms[b] were used for sequencing.	7,201 variants were detected after an initial filtering process to remove false-positive results. Of these, 2,633 were in protein-coding regions. Variant type (of 2,633 variants): Missense: 2,271 (86.3%) Silent: 754 (28.6%) Nonsense: 194 (7.4%) Splice site: 151 (5.7%) Indels: 67 (2.5%) Of missense variants, 397 (17.5%) were identified as affecting protein function using four different algorithms.	Follow-up studies based on results obtained from WES identified >20 genes with increased number of variants than expected by chance (P range, 2×10^{-14}–7×10^{-4}).
Xiong et al (2012)[34]	14 NSCLC patients from the People's Republic of China: three ADC; five SCC; and six BAC. Of 14 patients, 12 (85.7%) were smokers.	Tumors and matched normal adjacent lung tissue were used. Exome enrichment was performed using the Agilent SureSelect All Exome kit[a] to capture data from 212,919 exons and flanking regions. Illumina HiSeq 2000[b] was used for sequencing. Variants were validated using Sanger sequencing.	3,321 somatic variants were detected, with average of 237.2 per tumor (range, 11–963). Variant type (of 1,659 with potential biological function): Missense: 1,038 (62.6%) Silent: 447 (26.9%) Nonsense: 88 (5.3%) Splice site: 41 (2.5%) Insertions: 17 (1.0%) Deletions: 26 (1.6%) Of missense variants, 91 (8.8%) were identified as affecting protein function using four different algorithms.	Follow-up studies based on results obtained from WES identified >15 novel genes with increased number of variants than expected by chance (confirmed by Sanger sequencing). In addition, five previously identified genes were also detected.

Notes: Adapted with permission from Winifred S Hayes, Inc. Manufacturer details are as follows: [a]Agilent Technologies, Santa Clara, CA, USA; [b]Illumina Inc., San Diego, CA, USA.
Abbreviations: ADC, adenocarcinoma; BAC, bronchioloalveolar carcinoma; GA II, Genome Analyzer II; NSCLC, nonsmall cell lung carcinoma; SCC, squamous cell carcinoma; WES, whole exome sequencing; WGS, whole genome sequencing.

Clinical utility

Three studies evaluating treatment outcomes based on the results obtained through WES have suggested potential for clinical utility.[19,42,43]

- In one of the breast cancer studies described in Table 2, Ellis et al[19] (2012) evaluated the correlation between variants in specific genes (identified through WES and WGS) and response to aromatase inhibitors in patients with estrogen receptor-positive breast cancer. The protein Ki-67 is a marker of tumor proliferation, and high levels of this protein are associated with a poor prognosis even after aromatase inhibitor treatment. Genetic variants were analyzed at baseline and postsurgery in relation to Ki-67 levels. One gene, *GATA-binding protein 3* (*GATA3*), showed no correlation at baseline, but showed enrichment of variants in samples showing a decline in Ki-67 levels (*P*=0.01) postsurgery. These results suggest that the presence of *GATA3* variants may be a predictive marker to identify individuals who will respond to treatment with aromatase inhibitors.[19]

- A second study by Sohn et al[42] (2012) evaluated WES data from 174 patients with ovarian cancer who were part of the Cancer Genome Atlas study described in the "Clinical validity" section. In patients with high-grade serous ovarian cancer, the standard treatment is surgery followed by platinum-based chemotherapy. However, platinum resistance is observed in approximately 30% patients. Details of the study are:[42]
  - All patients had platinum response data and WES results available. From the WES results, individuals were classified into two groups, based on the number of variants detected: hypervariant (>61.9 variants per patient); or hypovariant (<61.9 per patient). In addition, platinum status after receiving prior platinum treatment was determined: platinum-sensitive (no platinum treatment for ≥6 months with no disease progression or recurrence); or platinum-resistant (<6 months to disease progression or recurrence).
  - Of the 174 patients, 122 (70.1%) and 52 (29.9%) were classified as platinum-sensitive and platinum-resistant, respectively. The number of variants was statistically significantly different between the two groups (mean number of variants of 67.2±38.9 and 49.4±25.2 for the sensitive and resistant groups, respectively; *P*<0.001). In addition, a multivariate logistic regression analysis identified somatic hypervariant as an independent risk factor for determining platinum response (odds ratio =3.616; 95% CI, 1.595–8.200; *P*=0.002).
  - Hypervariant status was also statistically significantly associated with longer overall survival (*P*<0.001) and progression-free survival (*P*=0.001).

- Finally, a third study evaluated WES in patients with acute lymphoblastic leukemia (ALL). ALL patients who experience remission following aggressive chemotherapy, but then relapse, are generally treated with chemotherapy again. However, 40% of these patients do not benefit from this subsequent chemotherapy treatment. Tzoneva et al[43] (2013) performed WES in five pediatric patients who had prior chemotherapy, remission, and relapsed ALL. Details of the study are:[43]
  - Variants were detected using the Agilent SureSelect kit (Agilent Technologies) for exome capture, and Illumina HiSeq 2000 (Illumina Inc.) for sequencing.
  - The WES analysis identified variants in the *5′-nucleotidase, cytosolic II (NT5C2)* gene. Follow-up analysis of this gene in 133 individuals with relapsed ALL identified 23 additional variants in this gene. Nine individuals with available DNA at time of diagnosis showed the absence of *NT5C2* variants at diagnosis, but these were acquired during relapse. In addition, individuals with *NT5C2* showed association with relapse even when receiving treatment (*P*=0.002), and early relapse compared to late relapse (*P*<0.05).
  - The presence of *NT5C2* variants in relapsed patients suggests that this may be a potential marker to identify individuals at increased risk for relapse, despite chemotherapy treatment.

Discussion and conclusion

An increasing number of WES studies have been performed in cancer patients in recent years, comparing tumor tissue to normal tissues. WES allows simultaneous detection of variants in protein-coding regions across the genome with a single assay instead of performing a large number of single tests to identify potential genetic contributors to the cancer being analyzed. WES also allows exploration of the heterogeneous nature of cancers. In this review, we evaluated the analytical validity, clinical validity, and clinical utility of WES for cancer indications.

Analytical validity evidence was provided by four studies – three studies in healthy individuals and one in cancer patients.[13–16] A number of limitations are noted. Two studies used data from only one individual, which limits statistical power.[13,15] In addition, only one study analyzed data

collected from cancer patients; however, cancer status was not specifically accounted for in the analyses.[14] The tissue types evaluated in cancer may be limited with respect to the quantity, preparation (for example, fresh frozen), and heterogeneity of clinically significant variants. In addition, the very wide spectrum of variants present in cancers may increase the complexity of data analysis required. Therefore, more studies that specifically address the analytical issues of WES in cancer specimens need to be conducted.

Furthermore, the exome comprises only about 1% of the genome and variants in noncoding regions, and other DNA alterations (including genomic rearrangements and copy number variations) will not be detected. Therefore, there is a possibility of missing clinically important variants. In addition, WES does not provide functional information about variants; consequently, results need to be followed up to determine the biological function of identified variants. Finally, WES relies on bioinformatic algorithms to infer variant information and the selection of genes for further analysis. Given that such algorithms are still being developed, current methods may be limited with respect to the optimal variant (and gene) selection.

Clinical validity studies have been mostly limited to assessment of genetic landscape information; whereby, details about the types of variants (including missense, nonsense, silent, insertions, and deletions) are provided. Follow-up of the WES results to identify genes with potential roles in cancer development have rarely been conducted, with no further assessment in most cases. Additional studies using adequate sample sizes and with follow-up of identified variants to assess their functional roles are therefore needed. Such analyses may lead to a better understanding of diagnostic and prognostic factors, and potential treatment avenues for cancer patients.

While no study provided direct evidence of clinical utility, three studies suggest the potential for clinical utility. Follow-up analyses showed enrichment of *GATA3* variants (identified by WES) in samples showing a decline in Ki-67 levels, which is a marker for response to aromatase inhibitor treatment. This association suggests that presence of *GATA3* variants may be a predictive marker to identify individuals who will respond to treatment with aromatase inhibitors.[19] A second study showed that somatic hypervariation detected through WES not only was a predictive factor for determining platinum-based chemotherapy response in ovarian cancer treatment, but it was also statistically significantly associated with longer

overall survival and progression-free survival.[42] Lastly, a WES study identified *NT5C2* variants that were associated with ALL relapse even when patients were receiving treatment, suggesting that *NT5C2* may be a potential marker to identify individuals who may experience ALL relapse despite chemotherapy treatment.[43] While these three studies suggest potential for clinical utility for WES in cancer, there are no studies that specifically demonstrate improvements in patient outcomes.

To assess the potential impact of studies published after June 2013, an updated search of the literature was conducted in March 2014. Although additional studies have been published since June 2013, these represented further genetic landscape evaluation with no additional evidence for clinical utility.

WES results may have ethical implications. One ethical issue that may arise with WES studies is the identification of an incidental finding with clinical significance, ie, a variant with a known disease-causing effect, but for a different condition than that being studied. This information may affect not only a patient, but it may have implications for the patient's family members. The methods of providing such information to patients are not entirely clear, although some recommendations have recently been made by the American College of Medical Genetics and Genomics.[44] In addition, given the limited number of follow-up studies (and small sample sizes) addressing treatment outcomes, there are potential safety concerns if a treatment decision is based solely on WES results.

In conclusion, while there is evidence for clinical validity, mainly for breast cancer, clinical utility evidence for WES is lacking for all types of cancer. Therefore, WES has not yet been proven to improve outcomes for patients with cancer.

Acknowledgments

Disclosure

The authors are employees at Winifred S Hayes, Inc., an independent health technology research and consulting company. None of the employees at this company has any

financial or personal interest in any of the technologies reviewed by Winifred S Hayes, Inc. No input on report content or conclusions is permitted by manufacturers. The authors report no other conflicts of interest in this work.

References

1. Shyr D, Liu Q. Next generation sequencing in cancer research and clinical application. *Biol Proced Online*. 2013;15(1):4.
2. Cancer.org [homepage on the Internet]. *Cancer Facts and Figures 2013*. Atlanta: American Cancer Society; 2014. Available from: http://www.cancer.org/acs/groups/content/@epidemiologysurveilance/documents/document/acspc-036845.pdf. Accessed March 19, 2014.
3. Bick D, Dimmock D. Whole exome and whole genome sequencing. *Curr Opin Pediatr*. 2011;23(6):594–600.
4. Yu Y, Wu BL, Wu J, Shen Y. Exome and whole-genome sequencing as clinical tests: a transformative practice in molecular diagnostics. *Clin Chem*. 2012;58(11):1507–1509.
5. Bamshad MJ, Ng SB, Bigham AW, et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet*. 2011;12(11):745–755.
6. Chakravarti A. Population genetics – making sense out of sequence. *Nat Genet*. 1999;21(Suppl 1):56–60.
7. Youn A, Simon R. Identifying cancer driver genes in tumor genome sequencing studies. *Bioinformatics*. 2011;27(2):175–181.
8. BCM.edu [homepage on the Internet]. Cancer Exome Sequencing. Houston: Baylor College of Medicine Medical Genetics Laboratories; 2014. Available from: https://www.bcm.edu/geneticlabs/test_detail.cfm?testcode=9610. Accessed March 19, 2014.
9. Personalgenome.com [homepage on the Internet]. Personal Genome Diagnostics; CancerComplete. Baltimore: Personal Genome Diagnostics; 2014. Available from: http://personalgenome.com/cancercomplete. Accessed March 19, 2014.
10. Teutsch SM, Bradley LA, Palomaki GE, et al; EGAPP Working Group. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: methods of the EGAPP Working Group. *Genet Med*. 2009;11(1):3–14.
11. Veenstra DL, Piper M, Haddow JE, et al. Improving the efficiency and relevance of evidence-based recommendations in the era of whole-genome sequencing: an EGAPP methods update. *Genet Med*. 2013;15(1):14–24.
12. Gudgeon JM, McClain MR, Palomaki GE, Williams MS. Rapid ACCE: experience with a rapid and structured approach for evaluating gene-based testing. *Genet Med*. 2007;9(7):473–478.
13. Asan, Xu Y, Jiang H, et al. Comprehensive comparison of three commercial human whole-exome capture platforms. *Genome Biol*. 2011;12(9):R95.
14. Guo Y, Long J, He J, et al. Exome sequencing generates high quality data in non-target regions. *BMC Genomics*. 2012;13:194.
15. Clark MJ, Chen R, Lam HY, et al. Performance comparison of exome DNA sequencing technologies. *Nat Biotechnol*. 2011;29(10):908–914.
16. Hedges DJ, Guettouche T, Yang S, et al. Comparison of three targeted enrichment strategies on the SOLiD sequencing platform. *PLoS One*. 2011;6(4):e18595.
17. Banerji S, Cibulskis K, Rangel-Escareno C, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature*. 2012;486(7403):405–409.
18. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490(7418):61–70.
19. Ellis M, Ding L, Shen D, et al. Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature*. 2012;486(7403):353–360.
20. Shah SP, Roth A, Goya R, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature*. 2012;486(7403):395–399.
21. Stephens PJ, Tarpey PS, Davies H, et al; Oslo Breast Cancer Consortium (OSBREAC). The landscape of cancer genes and mutational processes in breast cancer. *Nature*. 2012;486(7403):400–404.
22. Gracia-Aznarez FJ, Fernandez V, Pita G, et al. Whole exome sequencing suggests much of non-BRCA1/BRCA2 familial breast cancer is due to moderate and low penetrance susceptibility alleles. *PLoS One*. 2013;8(2):e55681.
23. Thompson ER, Doyle MA, Ryland GL, et al; kConFab. Exome sequencing identifies rare deleterious mutations in DNA repair genes FANCC and BLM as potential breast cancer susceptibility alleles. *PLoS Genet*. 2012;8(9):e1002894.
24. Quesada V, Conde L, Villamor N, et al. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat Genet*. 2011;44(1):47–52.
25. Wang L, Lawrence MS, Wan Y, et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N Engl J Med*. 2011;365(26):2497–2506.
26. Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med*. 2013;368(22):2059–2074.
27. Lohr JG, Stojanov P, Lawrence MS, et al. Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. *Proc Natl Acad Sci U S A*. 2012;109(10):3879–3884.
28. Zhang J, Grubor V, Love CL, et al. Genetic heterogeneity of diffuse large B-cell lymphoma. *Proc Natl Acad Sci U S A*. 2013;110(4):1398–1403.
29. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature*. 2011;474(7353):609–615.
30. Liang H, Cheung LW, Li J, et al. Whole-exome sequencing combined with functional genomics reveals novel candidate driver cancer genes in endometrial cancer. *Genome Res*. 2012;22(11):2120–2129.
31. Kinde I, Bettegowda C, Wang Y, et al. Evaluation of DNA from the Papanicolaou test to detect ovarian and endometrial cancers. *Sci Transl Med*. 2013;5(167):167ra4.
32. Zhao S, Choi M, Overton JD, et al. Landscape of somatic single-nucleotide and copy-number mutations in uterine serous carcinoma. *Proc Natl Acad Sci U S A*. 2013;110(8):2916–2921.
33. Liu P, Morrison C, Wang L, et al. Identification of somatic mutations in non-small cell lung carcinomas using whole-exome sequencing. *Carcinogenesis*. 2012;33(7):1270–1276.
34. Xiong D, Li G, Li K, et al. Exome sequencing identifies MXRA5 as a novel cancer gene frequently mutated in non-small cell lung carcinoma from Chinese patients. *Carcinogenesis*. 2012;33(9):1797–1805.
35. Imielinski M, Berger AH, Hammerman PS, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell*. 2012;150(6):1107–1120.
36. Seshagiri S, Stawiski EW, Durinck S, et al. Recurrent R-spondin fusions in colon cancer. *Nature*. 2012;488(7413):660–664.
37. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012;487(7407):330–337.
38. Lindberg J, Mills IG, Klevebring D, et al. The mitochondrial and autosomal mutation landscapes of prostate cancer. *Eur Urol*. 2013;63(4):702–708.
39. Barbieri CE, Baca SC, Lawrence MS, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat Genet*. 2012;44(6):685–689.
40. Stransky N, Egloff AM, Tward AD, et al. The mutational landscape of head and neck squamous cell carcinoma. *Science*. 2011;333(6046):1157–1160.
41. Agrawal N, Frederick MJ, Pickering CR, et al. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science*. 2011;333(6046):1154–1157.

42. Sohn I, Jung WY, Sung CO. Somatic hypermutation and outcomes of platinum based chemotherapy in patients with high grade serous ovarian cancer. *Gynecol Oncol*. 2012;126(1):103–108.

43. Tzoneva G, Perez-Garcia A, Carpenter Z, et al. Activating mutations in the NT5C2 nucleotidase gene drive chemotherapy resistance in relapsed ALL. *Nat Med*. 2013;19(3):368–371.

44. Green RC, Berg JS, Grody WW, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med*. 2013;15(7):565–574. Available from: http://www.acmg.net/docs/ACMG_Releases_Highly-Anticipated_ Recommendations_on_Incidental_Findings_in_Clinical_Exome_and_ Genome_Sequencing.pdf. Accessed March 19, 2014.