FULL PAPER

BJC

British Journal of Cancer (2015) 112, 1411–1420 | doi: 10.1038/bjc.2015.80

Keywords: massively parallel sequencing; diagnostics; FFPE; cancer genomics; actionable mutations; clinical seqeuncing

Assessing the clinical value of targeted massively parallel sequencing in a longitudinal, prospective population-based study of cancer patients

S Q Wong*[,1,2], A Fellowes[1], K Doig[3,4], J Ellul[3], T J Bosma[1], D Irwin[5], R Vedururu[1], A Y-C Tan[1], J Weiss[6], K S Chan[7], M Lucas[8], D M Thomas[2,4,9], A Dobrovic[1,2,6,10], J P Parisot[2,4] and S B Fox[1,2,4,10]

[1]Department of Pathology, Peter MacCallum Cancer Centre, East Melbourne, Victoria 3002, Australia; [2]Division of Cancer Research, Peter MacCallum Cancer Centre, St Andrews Place, East Melbourne, Victoria 3002, Australia; [3]Division of Cancer Research, Department of Bioinformatics, Peter MacCallum Cancer Centre, East Melbourne, Victoria 3002, Australia; [4]Sir Peter MacCallum Department of Oncology, The University of Melbourne, Parkville, Victoria 3010, Australia; [5]Agena Bioscience, Herston, Brisbane 4006, Australia; [6]Translational Genomics and Epigenomics Laboratory, Olivia Newton-John Cancer Research Institute, The Olivia Newton-John Cancer and Wellness Centre, Heidelberg, Victoria 3084, Australia; [7]Department of Pathology, Singapore General Hospital, Singapore 169608, Singapore; [8]Clinical Informatics and Data Management Unit, Alfred Centre, Monash University, Melbourne, Victoria 3004, Australia; [9]The Kinghorn Cancer Centre, Garvan Institute of Medical Research, Victoria Street, Darlinghurst, New South Wales 2010, Australia and [10]Department of Pathology, The University of Melbourne, Parkville, Victoria 3010, Australia

Introduction: Recent discoveries in cancer research have revealed a plethora of clinically actionable mutations that provide therapeutic, prognostic and predictive benefit to patients. The feasibility of screening mutations as part of the routine clinical care of patients remains relatively unexplored as the demonstration of massively parallel sequencing (MPS) of tumours in the general population is required to assess its value towards the health-care system.

Methods: Cancer 2015 study is a large-scale, prospective, multisite cohort of newly diagnosed cancer patients from Victoria, Australia with 1094 patients recruited. MPS was performed using the Illumina TruSeq Amplicon Cancer Panel.

Results: Overall, 854 patients were successfully sequenced for 48 common cancer genes. Accurate determination of clinically relevant mutations was possible including in less characterised cancer types; however, technical limitations including formalin-induced sequencing artefacts were uncovered. Applying strict filtering criteria, clinically relevant mutations were identified in 63% of patients, with 26% of patients displaying a mutation with therapeutic implications. A subset of patients was validated for canonical mutations using the Agena Bioscience MassARRAY system with 100% concordance. Whereas the prevalence of mutations was consistent with other institutionally based series for some tumour streams (breast carcinoma and colorectal adenocarcinoma), others were different (lung adenocarcinoma and head and neck squamous cell carcinoma), which has significant implications for health economic modelling of particular targeted agents. Actionable mutations in tumours not usually thought to harbour such genetic changes were also identified.

Conclusions: Reliable delivery of a diagnostic assay able to screen for a range of actionable mutations in this cohort was achieved, opening unexpected avenues for investigation and treatment of cancer patients.

*Correspondence: Dr SQ Wong; E-mail: stephen.wong@petermac.org

Received 16 December 2014; revised 29 January 2015; accepted 1 February 2015; published online 5 March 2015

The paradigm of personalised medicine is to define tumours from individual patients so as to maximise the clinical benefit of therapy, while minimising the likelihood that a patient will receive toxic, expensive and/or ineffective treatment. Underlying this paradigm is the classification of tumours on the basis of histological and/or molecular features. Tumour classification determines the diagnosis discussed with the patient and subsequently informs about surgical, radiotherapeutic and chemotherapeutic management. However, as many drug decisions are now based on a molecular target independent of the tissue of origin, there has been debate in the literature as to whether conventional histological classification remains useful (Swanton and Caldas, 2009; West *et al*, 2012). Targeting gene mutations within tumours rather than necessarily treating the particular tumour type is not only of great scientific interest but has profound and long-lasting consequences for the practise of cancer medicine, for how pathology is configured and for the design of clinical trials *en route* to drug approval.

Translating high-throughput sequencing technologies in the diagnostic detection of actionable mutations is developing as the next fundamental step in the management of cancer patients and is becoming the standard of practice for tumour specimens. However, there are several important factors that are critical for the integration of genomic findings into clinical practice.

From a demographic perspective, data from these large cancer genomic studies are not always representative of the general population and can be skewed if recruitment of patients depends heavily on advanced-stage disease to ensure adequate tissue is available for testing. While some projects embedded in the International Cancer Genome Consortium and the Cancer Genome Atlas (TCGA; Wood *et al*, 2007; Hudson *et al*, 2010; Imielinski *et al*, 2012; Shah *et al*, 2012; Kandoth *et al*, 2013) have specifically selected primary, pre-treatment tumours, the individual institutions are usually large tertiary or quaternary centres rather than hospitals with a majority of samples from these studies being fresh-frozen rather than clinical samples. Consequently, it is unknown whether the mutational spectrum seen in these large-scale studies will be mirrored in the general population of cancer patients. Genetic characterisation combined with clinical information will be needed to enable more accurate tumour classification for diagnosis, prognosis and treatment stratification. Importantly, despite the commitment of patients, the health-care profession, industry and government to this approach, there is a paucity of data as to whether personalised cancer therapy is an affordable or an efficient method to deliver care to cancer patients, particularly from a societal health perspective.

While there have been recent advances in genomic technologies, the use of formalin-derived DNA, integration of a high-throughput and sensitive sequencing workflow and incorporation of a variant management system are major considerations in the implementation of an effective workflow in a molecular diagnostic setting.

To establish the infrastructure and patient cohort able to answer these questions, we established the Cancer 2015 cohort study, a prospective, multi-institutional study of newly diagnosed cancer patients. The aims of the study were to determine (1) the limitations of the quantity and quality of DNA template required for sequencing, (2) the attrition rate of clinical samples submitted for these types of testing regimes, (3) the resources needed in the interpretation of the mutations called in a high-throughput clinical setting particularly those derived from low-frequency variants due to tumour heterogeneity or sequence artefacts and (4) the population-based frequency of actionable mutations to enable accurate health economic modelling with corresponding targeted and conventional therapies.

MATERIALS AND METHODS

Cancer patients. Cancer 2015 is a large-scale, prospective, longitudinal, multisite cohort study of incident-first cancers in the Victorian population (Parisot *et al*, in submission). Results presented in this paper represent phase 1 of the Cancer 2015 study aimed at establishing a feasible patient recruitment and molecular pathology workflow embedded in a diagnostic pathology laboratory. Patients were recruited from the Peter MacCallum Cancer Centre, Royal Melbourne Hospital, Cabrini Hospital, Geelong Hospital and Warrnambool Hospital, institutions representing a cancer centre, a major general hospital, a regional and a rural hospital. Formalin-fixed and paraffin-embedded (FFPE) tumour blocks or sections were acquired from anatomical pathology laboratories performing the diagnosis. Although not macrodissected, representative tumour-only samples were sent by referring pathologists to the central molecular pathology laboratory at the Peter MacCallum Cancer Centre for DNA extraction and profiling. A schematic of the workflow is shown in Supplementary Figure 1. This study was approved by the Human Research Ethics Committee at the Peter MacCallum Cancer Centre (HREC number 11/69) and all participating hospitals.

DNA extraction. Up to 10 sections of 5-μm thickness were cut from each block. Sections were stained with 0.5% methyl green to assist with scraping of cells from slides. The scraped cells then underwent proteinase K digestion overnight at 56 °C. DNA from FFPE sections were extracted using the DNeasy blood and tissue kit (Qiagen, Hilden, Germany) as per the manufacturer's instructions. DNA quantification was performed using the Qubit dsDNA HS Assay kit using the Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). Qubit readings were used as a guide for dilution of the DNA samples.

Amplicon cancer panel. The TruSeq Amplicon Cancer Panel (TSACP; Illumina, San Diego, CA, USA) comprises of 212 amplicons from 48 genes that are simultaneously amplified in a single-tube reaction (Supplementary Table 1 for full list of genes). A minimum of 50 ng of DNA was used for molecular profiling according to the manufacturer's instructions with the MiSeq system (Illumina) used for paired end sequencing with v1 or v2 150-bp kits.

Bioinformatic alignment and variant calling. CASAVA v1.8.2 was used to perform sample demultiplexing and to convert BCL files generated from the MiSeq instrument into FastQ files containing short-read data. Using the primer sequences that are present in the data, short reads were first assigned to their respective amplicon. A glocal alignment on the basis of a modified Needleman–Wunsch algorithm (Needleman and Wunsch, 1970) was then performed between the reads and the hg19-derived amplicon reference sequence. Sequence variants were detected using VarScan2 (Koboldt *et al*, 2012).

Variant curation and annotation. To be acceptable for variant analysis, a quality-control (QC) filter was established that required a sample to have a minimum of 150 000 total mapped reads ($\sim 750\times$ mean coverage per amplicon). Raw variants from these passed samples were deposited into 'Path-OS': an in-house web-based variant management system. To restrict our analysis to high confidence variants, only variants that had coverage of at least 100 total reads, greater than 50 variant reads and at least 8% variant frequency were further analysed. Variants recurring in more than 50% of the samples (representing likely sequencing artefacts) or variants with a high global allele frequency ($\geqslant 1.0\%$) in the 1000 genomes database were flagged and removed from this curated list. All variants were annotated using the Ensembl VEP (McLaren

et al, 2010) and Annovar software (Wang *et al*, 2010) to give inferred protein consequences and nomenclature.

Variants were stratified into one of five clinical relevance classes on the basis of their predictive, prognostic or diagnostic value (See Supplementary Methods for description of classification method).

Database comparisons. To compare the Cancer 2015 Cohort with the incidence of specific tumour types seen in the general Victorian population, the Victorian Cancer Registry (VCR) was contracted to provide de-identified data for all reported, incident-first diagnoses by cancer site (tumour type) from 2011. An additional survey by the VCR of 2845 randomly selected tumour-stratified cases from 2011, the Cancer 2015 Reference Cohort (RC), was used to determine biases in tumour stage in the Cancer 2015 data set. For a detailed version of the protocol used, see Supplementary Methods.

To allow a comparison of the prevalence of mutations in the Cancer 2015 cohort *vs* the Catalogue of Somatic Mutations in Cancer (COSMIC; Forbes *et al*, 2010) and TCGA databases, tumour streams were stratified into distinct histological subtypes. Tumour streams from the Cancer 2015 cohort with the largest number of patients were analysed including breast-invasive carcinoma, colorectal adenocarcinoma, lung adenocarcinoma and head and neck squamous cell carcinoma. Provisional mutation prevalence was extracted from TCGA data using the cBioportal resource (Cerami *et al*, 2012). TCGA genes that were also present in COSMIC with high mutational prevalence (using the cancer browser and top 20 genes display) were analysed.

Orthogonal validation. A subset of samples ($n = 74$) was orthogonally validated for variant calls produced by the TSACP. This consisted of different panels of the Agena Bioscience MassARRAY platform (OncoCarta v1, v2, v3, LungCarta, OncoFocus, PanCarta). Clinical samples that passed the sequencing QC filter and were represented on an Agena panel were tested. Agena Bioscience MassARRAY validation was performed as per the manufacturer's instructions.

Statistics. Fisher's exact tests were used to examine association between categorical variables. A nonparametric Spearman correlation was used to investigate associations between the TSACP allele frequency calls *vs* the allele frequency calls from the Agena Bioscience MassARRAY testing. All analyses were performed using the STATA software version 6.01 (College Station, TX, USA).

RESULTS

Sample quality-control and sequencing performance. The Cancer 2015 Cohort has reached 1094 patients with newly diagnosed cancers. Of these, 936 samples yielded DNA of sufficient quantity for sequencing using the TSACP (>50 ng). After sequencing, a QC filter was established for samples with insufficient read coverage

Table 1. Sequenced Cancer 2015 patients based on tumour stream, recruiting institution, gender and mutation rate

	Cabrini Institute		Geelong Hospital		Peter MacCallum Cancer Centre		Royal Melbourne Hospital		Warrnambool Hospital				
Gender	F	M	F	M	F	M	F	M	F	M	Number of patients	Number of mutations	Average mutation rate
Anal	0	0	2	2	1	0	0	0	0	0	5	8	1.60
Bladder	0	1	0	3	0	1	2	4	0	2	13	15	1.15
Bone and soft tissue	0	0	0	0	4	19	0	0	0	0	23	23	1.00
Breast	48	2	37	0	1	0	24	0	23	0	135	282	2.09
Cancers of unknown primary	0	1	2	1	8	3	0	0	3	0	18	109	6.06
Central nervous system	0	0	0	2	0	0	3	7	0	0	12	19	1.58
Cervical	1	0	0	0	24	1	0	0	1	0	27	33	1.22
Colorectal	31	31	7	11	0	3	1	0	11	6	101	359	3.55
Endometrial	11	0	0	0	1	0	0	0	0	0	12	24	2.00
Head and neck	0	0	1	7	16	74	7	10	0	0	115	325	2.83
Hepatic	0	1	0	1	0	1	0	0	0	1	4	4	1.00
Lung	7	8	4	5	8	10	5	9	1	2	59	200	3.39
Lymphoma	0	0	0	0	1	0	0	0	0	0	1	1	1.00
Melanoma	1	3	0	0	2	8	1	0	0	0	15	50	3.33
Oesophagogastric	0	2	4	3	0	4	4	3	0	2	22	52	2.36
Other	7	0	1	5	3	2	1	1	1	0	21	40	1.90
Ovarian	7	0	0	0	3	0	0	0	2	0	12	22	1.83
Pancreatic	1	3	1	2	0	0	1	2	0	0	10	10	1.00
Prostate	0	61	0	6	0	38	0	8	0	0	113	190	1.68
Renal	1	4	0	0	2	3	5	9	1	0	25	86	3.44
Testicular	0	1	0	1	0	3	0	1	0	0	6	6	1.00
Thyroid	0	0	0	0	3	1	0	0	0	0	4	3	0.75
Unknown	1	1	2	5	2	8	31	44	2	5	101	156	1.54
Total	116	119	61	54	79	180	85	99	45	18	854	2017	2.36

Abbreviations: F = females; M = males.

because of poor processing or DNA quality. In total, 854 samples (78%) passed sequencing QC filters (Supplementary Table 2) with the attrition summarised in Supplementary Figure 2. For samples that passed the QC metric filter, the mean coverage across all amplicons was 3900 × per sample with an average of 975 229 reads for each sample. Target efficiency was good with 92.3% of amplicon reads mapping to reference on average per sample. Amplicons for *CDKN2A* consistently had low read coverage because of high GC content and were subsequently removed from the downstream analysis.

Variant filtering and curation. Raw variant counts for the tumour samples analysed to date exceeded 100 000. A substantial proportion of these variants were below the 10% range and primarily C>T/G>A changes, consistent with formalin-induced sequence artefacts (Supplementary Figure 3A; Do *et al*, 2013; Wong *et al*, 2014). Variant peaks between the allele frequency ranges of 40–60 and 80–100% mostly represented single-nucleotide polymorphisms or panel-specific sequencing artefacts. Rule-based filtering using read coverage/allele thresholds, likelihood of being an artefact and global allele frequency resulted in a total of 2017 high confidence-curated variants from all 854 samples (Supplementary Table 3).

Validation of mutations. Orthogonal validation of variant calls was performed on 74 samples using Agena panels. In total, 91 variants were identified that had the capability to also be identified using an Agena panel. Concordance of variants detected by TSACP and validated on the Agena MassARRAY was 100%. However, two variants (12K0099-*NRAS* c.182A>G, p.Gln61Arg and 12K0377-*KRAS* c.37G>T, p.Gly13Cys) were detected by the Agena OncoFocus panel that were not present in the final curated list of TSACP variants (97.8% sensitivity). Current filtering thresholds on the basis of minimum variant read coverage and variant allele frequency precluded these from being included in the final curated list of variants with high confidence (Supplementary Table 4).

A comparison between the variant allele frequencies called from the TSACP and the Agena MassARRAY assay showed a strong positive correlation between the two platforms (Supplementary Figure 3B; $r = 0.928$, $P < 0.0001$).

Cancer 2015 cohort characteristics. The breakdown of samples that passed QC filtering across institutions and according to gender is shown in Table 1. Most participants were 50 years of age at diagnosis, with a median age of 63 ± 13 years. The number of participants was balanced by gender, with the exception of the Peter MacCallum Cancer Centre because of the high number of patient accruals for prostate and head and neck cancers from this centre, and the paucity of gynaecologic cancers.

To assess whether patients were a proper representation of those seen clinically, a comparison was performed between Cancer 2015 QC passed samples *vs* the incidence rates reported by the VCR (Figure 1A). Although generally proportionate to the VCR, there was a significant difference in the incidence of specific cancer types (Head and neck, cervical and melanoma) biased by the limited types of recruiting hospital sites ($P < 1.0 \times 10^{-4}$).

The stage groups of the Cancer 2015 patients were compared with a randomly sampled set using the VCR registry data from 2011 (Figure 1B). The Cancer 2015 cohort has a fairly even representation of each cancer staging group, but an under-representation of Stage I group cancers ($P < 1.0 \times 10^{-9}$). This may reflect some referral biases of more advanced cancers to the tertiary centres at which Cancer 2015 was open.

Classification of actionable mutations. A mean of 2.4 mutations per sample was found. Mutagen-induced tumours tended to have higher mutation rates. For example, melanomas contained on average 3.3 mutations per tumour, lung 3.4 mutations per tumour, whereas bone and soft tissue tumours contained one mutation per

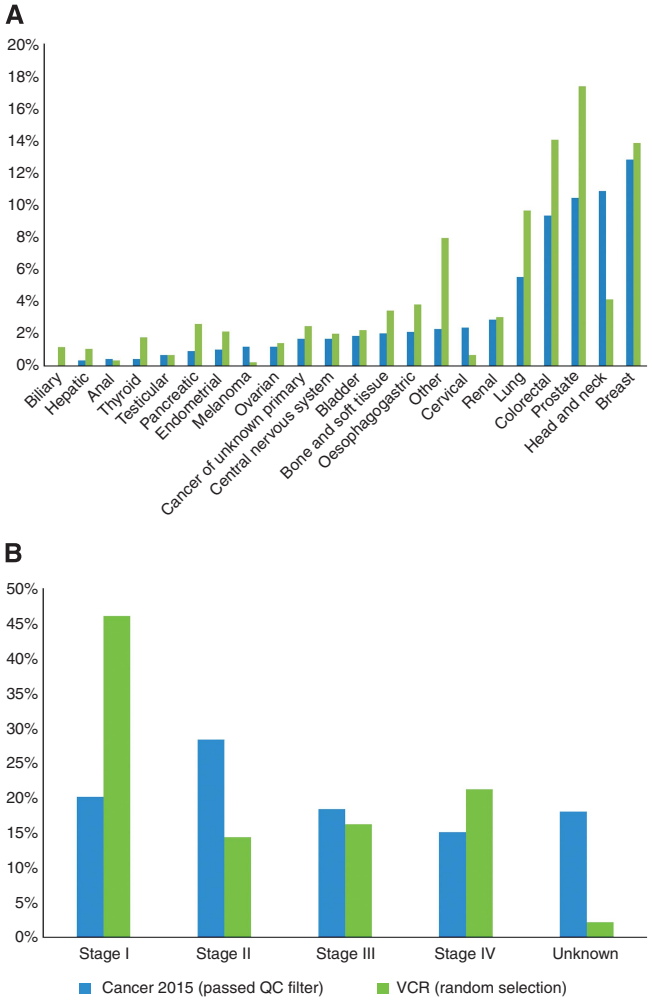

Figure 1. VCR Reference for ascertainment bias in the Cancer 2015 cohort. (**A**) The Cancer 2015 Cohort by cancer type, compared with the VCR 2011 census of solid-cancers only (with removal of paediatric and haematological cancers). Note: melanoma incidences represent advanced stages only. (**B**) The Reference Cohort obtained from the VCR 2011 census of solid-cancers data segmented into various cancer staging groups compared with the Cancer 2015 Cohort as a percentage of each random sample number.

tumour (Table 1). Interestingly, cancers of unknown primary (CUP) were commonly mutated (6.06 per tumour), suggesting that some of the specimens may have a mutagenic origin and may partly explain why these tumours are more refractory to treatment.

Classification of curated mutations was resolved around a stratification approach adapted by Wagle *et al* (2012) using the frequency of mutations according to gene, type of mutation and the type of actionable mutation (Figure 2A). Approximately 63% (534 out of 854) of patients had at least one clinically relevant mutation (Classes I–III). Overall, 31% of patients had a variant of prognostic/diagnostic significance (Class II), with 26% having a variant that provides sensitivity or resistant information to an approved or preclinical drug available in principle (Class IA/IB). However, there were a substantial number of patients (34%) who had a variant of unknown clinical significance (Class III).

As expected with this tumour set, *TP53* was the most mutated gene, whereas mutations in *NPM1* were not detected. Frameshift and nonsense mutations were more prevalent in the tumour-suppressor genes (TSGs) with few class I mutations present (Supplementary Table 5). Mutations were scattered through all TSGs with no defined hotspot except in canonical regions that

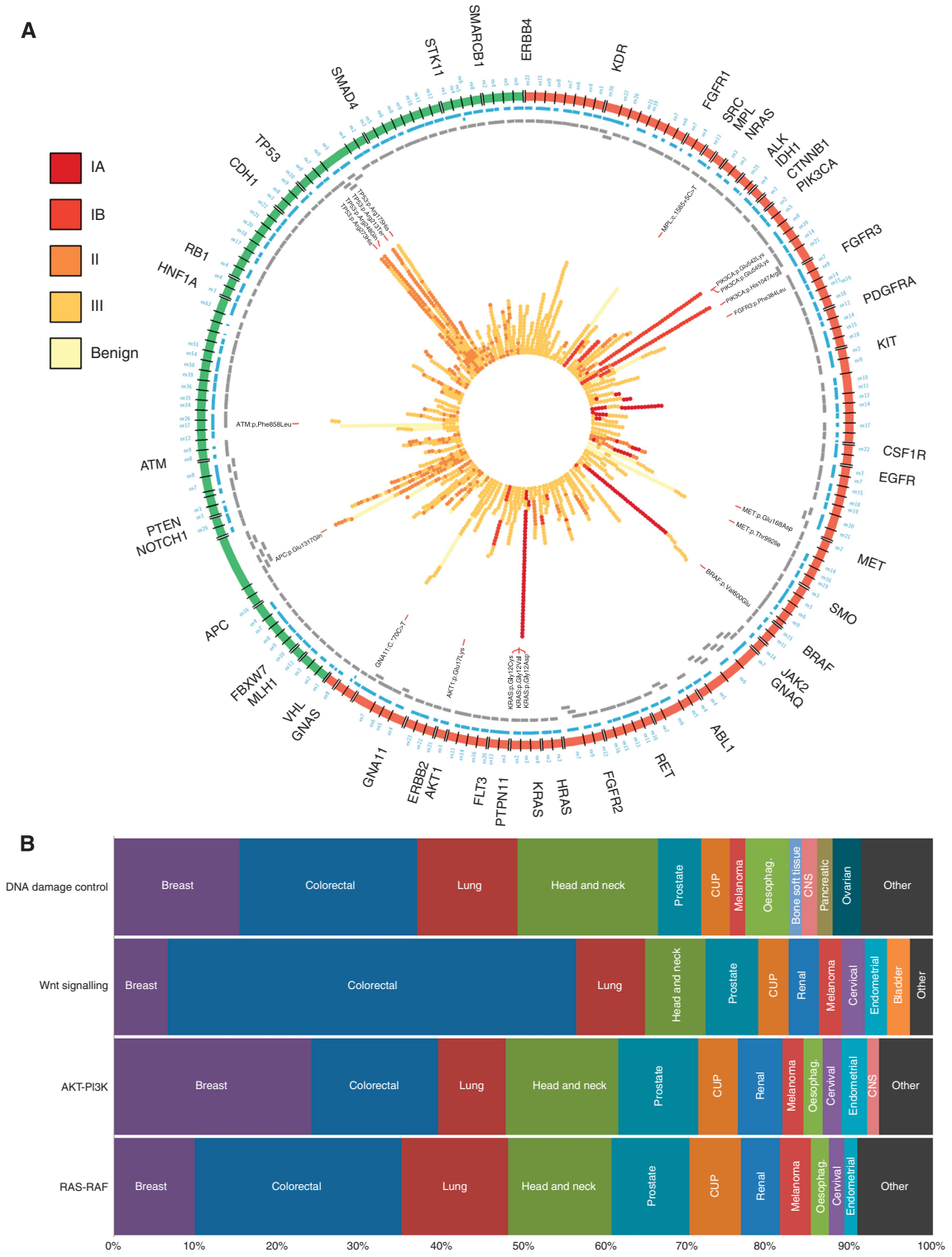

Figure 2. Mutational landscape of actionable mutations and pathways in the Cancer 2015 cohort. (**A**) Landscape of actionable mutations from the Cancer 2015 cohort. Tracks are (from outside in): Gene name, Exon label with type of cancer gene (green: tumour-suppressor gene, orange: oncogene), exon size shown as a blue tile, amplicon covered by the TSACP platform (grey tiles) and variants occurring >10 times in the filtered data. Variants are colour-coded based on the type of actionable mutation: (I) sensitive or resistant to an, approved drug/treatment (IA) or experimental drug/treatment (IB). (II) Provides prognostic or diagnostic information based on significant functional or clinically characterisation, (III) Unknown significance due to lack of biological/functional evidence or (IV) benign. Recurrent mutations are also highlighted. (**B**) Tumour classification by the actionable pathway. Variants from patients were stratified based on known associated pathways, detailed in Supplementary Table 1. The overall percentage of variants in any particular pathway is shown in the *x* axis. In some cases, a gene was associated with multiple pathways, for example, *NRAS* for PI3K-Akt and Ras-Raf pathways. In some cases, multiple genes were mutated in the same pathway. Multiple variants in the same gene from the same patient were only counted once. Only tumour streams with more than five patients mutated in a pathway are shown, with other cases combined to the other subset.

encode functionally important domains, for example, the DNA-binding domain of *TP53* (exons 4–9). In contrast, oncogenes had frequent hotspot missense Class I mutations that are known to have predictive value (for example, Codon 600 mutations in *BRAF*).

Separation of mutations on the basis of pathways revealed that most cancer types were represented into one of four major pathways (Figure 2B; DNA damage control, Wnt signalling, Akt-PI3K or RAS-RAF signalling). Whereas there was some expected over-representation of mutations in specific pathways (e.g. colorectal cancer in the Wnt signalling pathway), there were also some unexpected findings such as a classical activating mutation in *NRAS* in a head and neck tumour. Intriguingly for CUP, there was an even representation of mutations in all four pathways consistent with the heterogeneous nature of this tumour type. Overall, tumours can be classified by mutations into the key molecular pathways that can provide valuable predictive and prognostic benefit for patients.

Comparison of the incidence of mutations in the cancer 2015 cohort to other institutional-based studies. To determine an accurate representation of mutations in the general population of cancer patients, we compared the prevalence of mutations with the COSMIC and TCGA databases to those observed in this study (Figure 3). Examination of both colorectal adenocarcinoma and breast-invasive carcinoma cases found no significant difference in the prevalence of mutations for both tumour types across all data sets. There was, however, some noticeable differences in the invasive breast carcinoma group including the lack of *AKT1* mutations reported in the TCGA data set and the higher rate of *ATM* mutation in the Cancer 2015 data set.

In contrast, there was a difference in the mutation distribution for lung adenocarcinoma and head and neck squamous cell carcinoma when compared with respective TCGA data sets ($P = 0.0451$ and $P = 0.012$, respectively), but not with the COSMIC database. The under-representation of *EGFR* mutations in lung adenocarcinoma, but over-representation of *PIK3CA* mutations in head and neck carcinoma in the TCGA data set, is of particular interest as both genes have known predictive and prognostic value to patients. *TP53* mutations were almost twice as common in patients in the TCGA head and neck data set compared with the COSMIC and Cancer 2015 cohorts. Whereas there was no significant difference in the staging of lung cancer patients compared with the Cancer 2015 RC, there was a significant bias in the recruitment of stage IV tumours for head and neck cancers ($P < 1 \times 10^{-4}$).

DISCUSSION

The integration of massively parallel sequencing (MPS) into clinical practice using robust health economic modelling is an urgent task for health services worldwide. To address this issue, the Cancer 2015 study established a framework for a molecular pathology workflow to screen diagnostic tumour samples for common cancer genes. The feasibility of using MPS in a clinical setting has been demonstrated previously. However, previous studies were not reflective of cases seen in routine clinical practice, as they were either FFPE samples derived from cell line material (Tsongalis *et al*, 2013) or based on a selected number of cases from specific tumour streams (Singh *et al*, 2013; Wiesweg *et al*, 2013; Bourgon *et al*, 2014). Moreover, unlike the amplicon-based approach used in this study, many of the platforms utilised require much larger amounts of input DNA with turnaround times in excess of desirable practice (Wagle *et al*, 2012; Frampton *et al*, 2013; Pritchard *et al*, 2014).

Compared with other large-scale genomic projects, this study was aimed to be epidemiological in nature, as accurate assessment is critical in the establishment of an effective intervention strategy

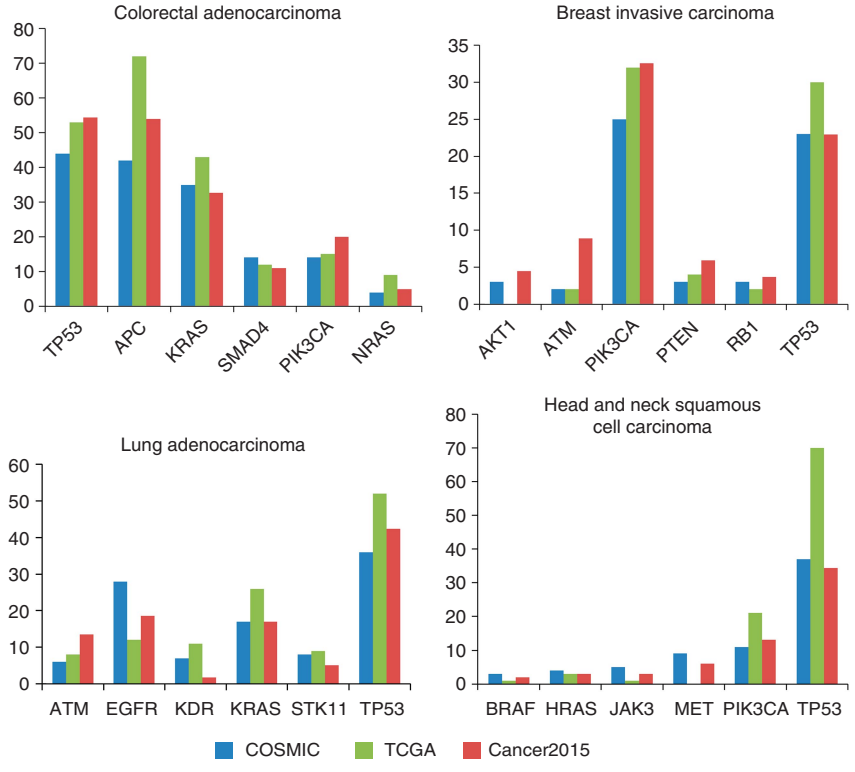

Figure 3. The prevalence of mutations in common cancer genes from the Cancer 2015 cohorts compared with other institutional-based series. Reported prevalence of mutations in colorectal adenocarinoma, breast-invasive carcinoma, lung adenocarcinoma and head and neck squamous cell carcinoma as reported by public mutational catalogues (COSMIC and TCGA) compared with the Cancer 2015 cohort.

for patient care including clinical trial design, types of genes to screen, genetic counselling, and so on. Population-based data are critical to understand not only the biology across the spectrum of any individual tumour type but also to assess the types and frequencies of mutations driving tumour progression through stages. Whereas this cohort under-represented stage I tumours, the major proportion of cancer patients suitable for intervention based on these genomic findings will have advanced disease. Nonetheless, this study was able to accurately capture the mutational incidence of some major tumour types in a cost- and time-effective manner. However, the data from this study have demonstrated that mutation frequencies can differ somewhat to those previously been reported. Notably, our data suggest that the mutation rate for *EGFR* is significantly higher in lung adenocarcinomas than anticipated from published data, suggesting that more patients could be eligible for EGFR inhibitors. This has major ramifications for regulatory agencies that regulate the safety, effectiveness and cost-effectiveness of drug treatments.

Given the significant bias in the recruitment of stage IV head and neck cancers in the Cancer 2015 cohort, we cannot conclude that *PIK3CA* mutations are under-represented in the general Victorian cancer population. This, however, provides valuable information from a health economics prospective. Head and neck cancer patients represent the highest average cost of any tumour stream due to ongoing radiotherapy treatment (on the basis of the national medical expenditure statistics). Findings from this study may lead to a more effective treatment strategy with many late-stage head and neck cancer patients potentially integrated into clinical trials for *PIK3CA* inhibitors (Janku *et al*, 2012) from the initial sites of recruitment.

Many therapeutic gene targets and pathways were found to be commonly mutated across different tumour streams. Interestingly, our findings are consistent with a previous study on a selection of CUP samples showing many somatic pathways activated in this tumour type (Tothill *et al*, 2013). The heterogeneous nature of actionable mutations in different pathways for CUPs gives a rationale for mutational profiling using a broad panel of genes.

To make progress towards the real-time reporting of clinical patients, some limitations have been accepted as a result of this study. First, there were a sizeable number of samples that were not successfully sequenced. Whereas more tumour materials could be made available at biopsy for testing, this is quite difficult in some circumstances, particularly from clinically challenging sites that give limited amounts of material. Resorting to orthogonal methodologies or newer technologies that have lower input requirements would be a suitable alternative strategy. Moreover, increasing the quality of input DNA could provide a higher level of successfully sequenced samples. This could be achieved by refining and standardizing fixation procedures, optimising storage conditions of tissue blocks and/or using alternative fixatives that minimise DNA damage (Do and Dobrovic, 2015). Importantly, upfront QC steps that assess the quality and quantity of DNA material will be paramount in directing a sample to its maximal screening potential.

Table 2. Recommendations in the processing and genomic testing of cancer specimens for mutational analysis and interpretation

Process step	Issue	Recommendation
Sample input	-Some MPS applications require large amounts of input DNA	-Efficient and high-throughput extraction methods are recommended (automation of extraction is suggested for tracking large numbers of samples) -Low elution volumes are also recommended to maximise DNA input -Standardized fixation methods and optimised storage conditions of tissue blocks that maximise the quality and quality of DNA extracted should be employed
Sample quality control	-FFPE-derived DNA is often fragmented, limiting the amount of useable material for MPS	-Integration of a quality-control step that assesses DNA integrity before sequencing -Use of auxiliary testing methods for samples that fail suitability for MPS
Sequencing platform	-MPS platforms can range widely in sequencing data output, processing times, running costs	-Currently, benchtop sequencers are best suited for diagnostic purposes because of ease of use, manageable data outputs, quicker processing times and lower running costs
Sequencing panel/assay	-Mutational profiling using MPS can range from a small panel of genes to whole exome/genome scale sequencing	-A small to medium panel of genes is generally preferred as it targets valuable sequence coverage to clinically informative genes rather than genes of low clinical value
Bioinformatics processing of sequencing data	-MPS can generate immense amounts of sequencing data -Raw sequencing data require multiple processing steps to generate variant calls	-Adequate data storage based on local or cloud-based systems -Automated and integrated bioinformatics pipeline dedicated to generate variants
Variant filtering	-System noise, technical artefacts and rare SNPs can make detection of somatic mutations difficult -Variants called from FFPE-derived DNA often display sequencing artefacts	-Rule-based filtering of variants should be applied to ensure that only high confidence variants are analysed -All actionable mutations should be validated internally though replication or/and through orthogonal testing
Interpretation	-Variants of unknown biological or clinical relevance can often be identified	-Information based on known variant prediction analysis or literature-based/database evidence can aid in the interpretation of variants -Multidisciplinary discussions in the interpretation of variants allowing a comprehensive and efficient approach in clinical management
Reporting	-The number of variants produced from MPS data make it difficult to decide what to report to a clinician	-A concise report that describes variants of most clinical applicability and that provides decision support should be produced -Comprehensive details of other relevant variants can be included supplementary to the main report
Workflow management	-MPS dramatically increases the number of samples tested -Owing to multiple loci tested, multiple mutations have to be analysed	-Incorporation of automation and a LIMS to streamline processes and shorten turnaround times -Implementation of a variant management system to catalogue mutations

Abbreviations: FFPE = Formalin-fixed and paraffin-embedded; LIMS = laboratory information management system; MPS = massively parallel sequencing; SNP = single-nucleotide polymorphism.

Complete validation of all actionable mutations was not performed in this study. However, replicates of the same sample would be a suitable approach to validate the detection of variants as well as mitigate sequencing errors. This assay was able to achieve a relatively good sensitivity compared with conventional sequencing methods. Although known pathogenic variants in the range 1–8% were observable, technical noise (contributed to by sequencing errors, artefacts of fixation and PCR artefacts) significantly obscured actual variants at this level, making detection less than 8% unrealistic at this time. Given the massive number of raw variants identified in this study, we cannot preclude that some curated variants may be false positives due to formalin-induced modifications. This also underpins the importance of a variant management system that is able to automatically and systematically filter variants with high confidence in order to ensure a high standard of reporting to clinicians.

Normal germline DNA was not sequenced in this study because of the increased sequencing cost for each case. Owing to the limited size of the regions screened in this study, common single-nucleotide polymorphism (SNP) variants were manageably identified through our variant management system and SNP data repositories. As gene panel sizes increase in the near future, particularly for targeted-capture or whole-genome sequencing platforms, this will undoubtedly be a major consideration in molecular pathology testing particularly in terms of costs, bioinformatic algorithms employed and incidental findings that may occur in sequencing germline DNA. SNP repositories such as the 1000 genomes project (Abecasis *et al*, 2012) and exome variant server could potentially provide a suitable means to subtract normal SNPs where normal tissue is unavailable or where sequencing of normal DNA is cost-prohibitive.

One of the challenges identified in this study is bridging the gap between research and clinical diagnostics for the interpretation of variants. The clinical significance of variant of unknown significance will undoubtedly be an area requiring more investigation in the future. Computational modelling and prediction will aid in the interpretation of these unknown variants. It is also feasible that both *in vitro* and *in vivo* models (Quintana *et al*, 2012) that functionally test the biological nature of mutations could be incorporated into diagnostic workflows, although given the need for a rapid turnaround time, this would be challenging. Ultimately, the interpretation of sequencing results to guide diagnosis and treatment decisions will likely require multidisciplinary expertise to unravel the functional and clinical implications of a result. Several of these and other recommendations should be considered for successful implementation of MPS for genomic testing of cancer patients (Table 2). The incorporation of automation and a laboratory information management system would streamline processes and could achieve a turnaround time of less than 2 weeks from sample acquisition to issuing a report for this type of platform.

CONCLUSIONS

The MPS technology enabled screening of multiple gene loci for mutations in a single workflow that is feasible across a wide range of tumour types. Our findings indicate that platforms that efficiently make use of DNA template without sacrificing analytical sensitivity or specificity will be more suited for a genomic testing in cancer patients. In the future, the molecular data from this cohort and other substudies will serve as a valuable resource in answering important questions on the makeup of cancers including their underlying aetiology, cancer biology regarding cell hierarchy and tumour progression, and help in defining patterns of response and mechanisms of resistance. The increased complexity of molecular profiling, issues with variant calls and their interpretation, and the paradigm shift whereby targeted therapies may no longer be tumour stream-specific, will result in a new model of triaging patients for appropriate molecularly guided therapies.

ACKNOWLEDGEMENTS

We gratefully acknowledge the cooperation of the following Victorian institutions: The Andrew Love Cancer Centre; Geelong Hospital, Barwon Health; Warrnambool Hospital; Southwest Health; The Peter MacCallum Cancer Centre; Ludwig Institute for Cancer Research, Austin Health; Royal Melbourne Hospital, Melbourne Health; Centre for Health Economics, Monash University; Department of Epidemiology and Preventative Medicine, The Alfred Centre, Monash University; Cabrini Health; Department of Pathology, University of Melbourne and Monash Institute of Medical Research. We thank all the cancer patients who participated in the study. We also thank all participating pathology laboratories involved in this study. We thank Helen Farrugia from the VCR for providing VCR data for this study. The results shown here are in whole or part based on data generated by the TCGA Research Network: http://cancergenome.nih.gov/. This study was supported by the Victorian Government through the Victorian Cancer Agency Translational Research Program, the National Collaborative Research Infrastructure Strategy (NCRIS) of the Australian Government and Therapeutic Innovation Australia. SQW is supported by the Melbourne Melanoma Project funded by the Victorian Cancer Agency Translational research program, established through the support of the Victor Smorgon Charitable Fund.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**(7422): 56–65.

Bourgon R, Lu S, Yan Y, Lackner MR, Wang W, Weigman V, Wang D, Guan Y, Ryner L, Koeppen H, Patel R, Hampton GM, Amler LC, Wang Y (2014) High-throughput detection of clinically relevant mutations in archived tumor samples by multiplexed PCR and next generation sequencing. *Clin Cancer Res* **20**(8): 2080–2091.

Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**(5): 401–404.

Do H, Dobrovic A (2015) Sequence artifacts in DNA from formalin-fixed tissues: causes and strategies for minimization. *Clin Chem* **61**(1): 64–71.

Do H, Wong SQ, Li J, Dobrovic A (2013) Reducing sequence artifacts in amplicon-based massively parallel sequencing of formalin-fixed paraffin-embedded DNA by enzymatic depletion of uracil-containing templates. *Clin Chem* **59**(9): 1376–1383.

Forbes SA, Tang G, Bindal N, Bamford S, Dawson E, Cole C, Kok CY, Jia M, Ewing R, Menzies A, Teague JW, Stratton MR, Futreal PA (2010) COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. *Nucleic Acids Res* **38**(Database issue): D652–D657.

Frampton GM, Fichtenholtz A, Otto GA, Wang K, Downing SR, He J, Schnall-Levin M, White J, Sanford EM, An P, Sun J, Juhn F, Brennan K, Iwanik K, Maillet A, Buell J, White E, Zhao M, Balasubramanian S, Terzic S, Richards T, Banning V, Garcia L, Mahoney K, Zwirko Z, Donahue A,

Beltran H, Mosquera JM, Rubin MA, Dogan S, Hedvat CV, Berger MF, Pusztai L, Lechner M, Boshoff C, Jarosz M, Vietz C, Parker A, Miller VA, Ross JS, Curran J, Cronin MT, Stephens PJ, Lipson D, Yelensky R (2013) Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat Biotechnol* **31**(11): 1023–1031.

Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, Bernabe RR, Bhan MK, Calvo F, Eerola I, Gerhard DS, Guttmacher A, Guyer M, Hemsley FM, Jennings JL, Kerr D, Klatt P, Kolar P, Kusada J, Lane DP, Laplace F, Youyong L, Nettekoven G, Ozenberger B, Peterson J, Rao TS, Remacle J, Schafer AJ, Shibata T, Stratton MR, Vockley JG, Watanabe K, Yang H, Yuen MM, Knoppers BM, Bobrow M, Cambon-Thomsen A, Dressler LG, Dyke SO, Joly Y, Kato K, Kennedy KL, Nicolas P, Parker MJ, Rial-Sebbag E, Romeo-Casabona CM, Shaw KM, Wallace S, Wiesner GL, Zeps N, Lichter P, Biankin AV, Chabannon C, Chin L, Clement B, de Alava E, Degos F, Ferguson ML, Geary P, Hayes DN, Johns AL, Kasprzyk A, Nakagawa H, Penny R, Piris MA, Sarin R, Scarpa A, van de Vijver M, Futreal PA, Aburatani H, Bayes M, Botwell DD, Campbell PJ, Estivill X, Grimmond SM, Gut I, Hirst M, Lopez-Otin C, Majumder P, Marra M, McPherson JD, Ning Z, Puente XS, Ruan Y, Stunnenberg HG, Swerdlow H, Velculescu VE, Wilson RK, Xue HH, Yang L, Spellman PT, Bader GD, Boutros PC, Flicek P, Getz G, Guigo R, Guo G, Haussler D, Heath S, Hubbard TJ, Jiang T, Jones SM, Li Q, Lopez-Bigas N, Luo R, Muthuswamy L, Ouellette BF, Pearson JV, Quesada V, Raphael BJ, Sander C, Speed TP, Stein LD, Stuart JM, Teague JW, Totoki Y, Tsunoda T, Valencia A, Wheeler DA, Wu H, Zhao S, Zhou G, Lathrop M, Thomas G, Yoshida T, Axton M, Gunter C, Miller LJ, Zhang J, Haider SA, Wang J, Yung CK, Cros A, Liang Y, Gnaneshan S, Guberman J, Hsu J, Chalmers DR, Hasel KW, Kaan TS, Lowrance WW, Masui T, Rodriguez LL, Vergely C, Bowtell DD, Cloonan N, deFazio A, Eshleman JR, Etemadmoghadam D, Gardiner BB, Kench JG, Sutherland RL, Tempero MA, Waddell NJ, Wilson PJ, Gallinger S, Tsao MS, Shaw PA, Petersen GM, Mukhopadhyay D, DePinho RA, Thayer S, Shazand K, Beck T, Sam M, Timms L, Ballin V, Lu Y, Ji J, Zhang X, Chen F, Hu X, Yang Q, Tian G, Zhang L, Xing X, Li X, Zhu Z, Yu Y, Yu J, Tost J, Brennan P, Holcatova I, Zaridze D, Brazma A, Egevard L, Prokhortchouk E, Banks RE, Uhlen M, Viksna J, Ponten F, Skryabin K, Birney E, Borg A, Borresen-Dale AL, Caldas C, Foekens JA, Martin S, Reis-Filho JS, Richardson AL, Sotiriou C, Thoms G, van't Veer L, Birnbaum D, Blanche H, Boucher P, Boyault S, Masson-Jacquemier JD, Pauporte I, Pivot X, Vincent-Salomon A, Tabone E, Theillet C, Treilleux I, Bioulac-Sage P, Decaens T, Franco D, Gut M, Samuel D, Zucman-Rossi J, Eils R, Brors B, Korbel JO, Korshunov A, Landgraf P, Lehrach H, Pfister S, Radlwimmer B, Reifenberger G, Taylor MD, von Kalle C, Majumder PP, Pederzoli P, Lawlor RA, Delledonne M, Bardelli A, Gress T, Klimstra D, Zamboni G, Nakamura Y, Miyano S, Fujimoto A, Campo E, de Sanjose S, Montserrat E, Gonzalez-Diaz M, Jares P, Himmelbauer H, Bea S, Aparicio S, Easton DF, Collins FS, Compton CC, Lander ES, Burke W, Green AR, Hamilton SR, Kallioniemi OP, Ley TJ, Liu ET, Wainwright BJ (2010) International network of cancer genome projects. *Nature* **464**(7291): 993–998.

Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, Cho J, Suh J, Capelletti M, Sivachenko A, Sougnez C, Auclair D, Lawrence MS, Stojanov P, Cibulskis K, Choi K, de Waal L, Sharifnia T, Brooks A, Greulich H, Banerji S, Zander T, Seidel D, Leenders F, Ansen S, Ludwig C, Engel-Riedel W, Stoelben E, Wolf J, Goparju C, Thompson K, Winckler W, Kwiatkowski D, Johnson BE, Janne PA, Miller VA, Pao W, Travis WD, Pass HI, Gabriel SB, Lander ES, Thomas RK, Garraway LA, Getz G, Meyerson M (2012) Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* **150**(6): 1107–1120.

Janku F, Wheler JJ, Westin SN, Moulder SL, Naing A, Tsimberidou AM, Fu S, Falchook GS, Hong DS, Garrido-Laguna I, Luthra R, Lee JJ, Lu KH, Kurzrock R (2012) PI3K/AKT/mTOR inhibitors in patients with breast and gynecologic malignancies harboring PIK3CA mutations. *J Clin Oncol* **30**(8): 777–782.

Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF, Wyczalkowski MA, Leiserson MD, Miller CA, Welch JS, Walter MJ, Wendl MC, Ley TJ, Wilson RK, Raphael BJ, Ding L (2013) Mutational landscape and significance across 12 major cancer types. *Nature* **502**(7471): 333–339.

Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* **22**(3): 568–576.

McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F (2010) Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**(16): 2069–2070.

Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* **48**(3): 443–453.

Pritchard CC, Salipante SJ, Koehler K, Smith C, Scroggins S, Wood B, Wu D, Lee MK, Dintzis S, Adey A, Liu Y, Eaton KD, Martins R, Stricker K, Margolin KA, Hoffman N, Churpek JE, Tait JF, King MC, Walsh T (2014) Validation and implementation of targeted capture and sequencing for the detection of actionable mutation, copy number variation, and gene rearrangement in clinical cancer specimens. *J Mol Diagn* **16**(1): 56–67.

Quintana E, Piskounova E, Shackleton M, Weinberg D, Eskiocak U, Fullen DR, Johnson TM, Morrison SJ (2012) Human melanoma metastasis in NSG mice correlates with clinical outcome in patients. *Sci Transl Med* **4**(159): 159ra149.

Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, Turashvili G, Ding J, Tse K, Haffari G, Bashashati A, Prentice LM, Khattra J, Burleigh A, Yap D, Bernard V, McPherson A, Shumansky K, Crisan A, Giuliany R, Heravi-Moussavi A, Rosner J, Lai D, Birol I, Varhol R, Tam A, Dhalla N, Zeng T, Ma K, Chan SK, Griffith M, Moradian A, Cheng SW, Morin GB, Watson P, Gelmon K, Chia S, Chin SF, Curtis C, Rueda OM, Pharoah PD, Damaraju S, Mackey J, Hoon K, Harkins T, Tadigotla V, Sigaroudinia M, Gascard P, Tlsty T, Costello JF, Meyer IM, Eaves CJ, Wasserman WW, Jones S, Huntsman D, Hirst M, Caldas C, Marra MA, Aparicio S (2012) The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* **486**(7403): 395–399.

Singh RR, Patel KP, Routbort MJ, Reddy NG, Barkoh BA, Handal B, Kanagal-Shamanna R, Greaves WO, Medeiros LJ, Aldape KD, Luthra R (2013) Clinical validation of a next-generation sequencing screen for mutational hotspots in 46 cancer-related genes. *J Mol Diagn* **15**(5): 607–622.

Swanton C, Caldas C (2009) Molecular classification of solid tumours: towards pathway-driven therapeutics. *Br J Cancer* **100**(10): 1517–1522.

Tothill RW, Li J, Mileshkin L, Doig K, Siganakis T, Cowin P, Fellowes A, Semple T, Fox S, Byron K, Kowalczyk A, Thomas D, Schofield P, Bowtell DD (2013) Massively-parallel sequencing assists the diagnosis and guided treatment of cancers of unknown primary. *J Pathol* **231**(4): 413–423.

Tsongalis GJ, Peterson JD, de Abreu FB, Tunkey CD, Gallagher TL, Strausbaugh LD, Wells WA, Amos CI (2013) Routine use of the Ion Torrent AmpliSeq Cancer Hotspot Panel for identification of clinically actionable somatic mutations. *Clin Chem Lab Med* **52**(5): 707–714.

Wagle N, Berger MF, Davis MJ, Blumenstiel B, Defelice M, Pochanard P, Ducar M, Van Hummelen P, Macconaill LE, Hahn WC, Meyerson M, Gabriel SB, Garraway LA (2012) High-throughput detection of actionable genomic alterations in clinical tumor samples by targeted, massively parallel sequencing. *Cancer Discov* **2**(1): 82–93.

Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**(16): e164.

West L, Vidwans SJ, Campbell NP, Shrager J, Simon GR, Bueno R, Dennis PA, Otterson GA, Salgia R (2012) A novel classification of lung cancer into molecular subtypes. *PLoS ONE* **7**(2): e31906.

Wiesweg M, Ting S, Reis H, Worm K, Kasper S, Tewes M, Welt A, Richly H, Meiler J, Bauer S, Hense J, Gauler TC, Kohler J, Eberhardt WE, Darwiche K, Freitag L, Stamatis G, Breitenbucher F, Wohlschlaeger J, Theegarten D, Derks C, Cortes-Incio D, Linden G, Skottky S, Lutkes P, Dechene A, Paul A, Markus P, Schmid KW, Schuler M (2013) Feasibility of preemptive biomarker profiling for personalised early clinical drug development at a Comprehensive Cancer Center. *Eur J Cancer* **49**(15): 3076–3082.

Wong SQ, Li J, Tan AY, Vedururu R, Pang JM, Do H, Ellul J, Doig K, Bell A, MacArthur GA, Fox SB, Thomas DM, Fellowes A, Parisot JP, Dobrovic A (2014) Sequence artefacts in a prospective series of formalin-fixed tumours tested for mutations in hotspot regions by massively parallel sequencing. *BMC Med Genomics* **7**(1): 23.

Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ, Shen D, Boca SM, Barber T, Ptak J, Silliman N, Szabo S, Dezso Z, Ustyanksky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kaminker JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Willson JK, Sukumar S, Polyak K, Park BH,

Pethiyagoda CL, Pant PV, Ballinger DG, Sparks AB, Hartigan J, Smith DR, Suh E, Papadopoulos N, Buckhaults P, Markowitz SD, Parmigiani G, Kinzler KW, Velculescu VE, Vogelstein B (2007) The genomic landscapes of human breast and colorectal cancers. *Science* **318**(5853): 1108–1113.

Supplementary Information accompanies this paper on British Journal of Cancer website (http://www.nature.com/bjc)