Sequencing structural variants in cancer for precision therapeutics

Geoff Macintyre[1], Bauke Ylstra[2], James D. Brenton[1,*]

1. Cancer Research UK Cambridge Institute, University of Cambridge, United Kingdom
2. Department of Pathology, VU University Medical Center, PO Box 7057, 1007 MB Amsterdam, The Netherlands

*Correspondence: james.brenton@cruk.cam.ac.uk (J.D. Brenton).

Keywords: shallow whole-genome sequencing, structural variation, clinical sequencing, cancer

Identification of mutations that guide therapy selection for patients with cancer is now routine in many clinical centres. The majority of assays used for solid tumour profiling use DNA sequencing to interrogate somatic point mutations as they are relatively easy to identify and interpret. However many cancers, including high-grade serous ovarian, oesophageal and small cell lung cancer, are driven by somatic structural variants that are not measured by these assays. Therefore, there is currently an unmet need for clinical assays that can cheaply and rapidly profile structural variants in solid tumours. In this review we survey the landscape of "actionable" structural variants in cancer and identify promising detection strategies based on massively parallel sequencing.

The necessity for increased profiling of structural variants in the clinic

Sequencing of tumour cohorts has provided key insights into tumour etiology[1,2] and has facilitated the development of new DNA based biomarkers for use in the clinic[3]. The current paradigm for biomarker discovery involves comprehensive, high-cost sequencing (usually exome or whole-genome) across moderately sized cohorts, then subsequent application of cheaper, targeted sequencing for biomarker detection in the clinic. This approach has yielded biomarkers that assist with diagnosis[4,5], prognosis[6,7], personalised therapy selection[8,9] and clinical trial stratification[10]. These biomarkers are now being internationally accepted into molecular pathology as shown by a recent landmark consensus whitepaper for nervous system tumour diagnosis[11] which will be incorporated into the World Health Organization classification of glioma.

Comprehensive approaches such as whole-genome (WGS) and whole-exome (WES) sequencing yield information on somatic point mutations, also known as single nucleotide variants (SNVs, See Glossary), small somatic insertions/deletions (INDELs), and somatic structural variants (SVs). These approaches have the benefit of identifying multiple mutations simultaneously, albeit at increased cost. However, given the reality that most actionable mutations are SNVs, the majority of clinical sequencing is currently performed using low-cost targeted gene panels that interrogate SNVs[12].

There are, however, many patients that will not benefit from SNV assays as their cancers are characterised by high numbers of SVs rather than SNVs. Furthermore, SV driven tumours may have few, if any, SNV drivers, as demonstrated by a recent large-scale pan-cancer study[13]. Ciriello and colleagues analysis of 3,299 tumours showed a hyperbolic relationship between the frequencies of SNVs and SVs across different tumour types, potentially arising from differences in aberrant DNA repair and mutagen exposure. The clinical implication of this work is that in the majority of cancers the major driver mutations are either SNV or SV mutations. For common cancers that are SV driven, such as high-grade serous ovarian[14], oesophageal[15], neuroblastoma[16], small-cell lung cancer[17], and triple-negative breast cancers[18], the current bias for implementation of SNV based assays leaves

clinicians with very limited data for precision medicine. In addition, these tumours are also enriched for loss of classical tumour suppressor genes including *TP53, RB1, NF1,* which are not directly actionable. There is a significant opportunity to extend clinical options for patients if cheaper methods for structural variant profiling can be developed for wider implementation, as exemplified by the repurposing of trastuzumab to treat gastric and other cancers with amplification of ERBB2[19].

Current state of the art in the clinic

Traditional cytogenetic techniques such as fluorescence *in situ* hybridisation (FISH) have been used clinically to profile SVs in haematological cancers for over 20 years[20]. However, when applied to solid malignancies, these technologies suffer from a number of drawbacks, the most significant being poor performance when applied to archival samples and, in contrast to haematological cancer, significant barriers to their use on disaggregated tumour tissue in vitro culture or from flow cytometry analyses. Recent European Guidelines for cytogenetic investigations in tumours have now recommended the use of orthogonal technologies to improve SV detection[21]. Alternative technologies such as array comparative genomic hybridization (aCGH) have been used in large-scale clinical trials to select patients with specific SVs using the OncoCopy[10] and OncoArray[22] platforms. While these studies show that aCGH is a viable clinical assay, the main weaknesses are a lack of standardised bioinformatics pipelines[23], high sample costs compared to equivalent sequencing technologies, an inability to interrogate balanced SVs, and the requirement for a different workflow in addition to those in place for panel-based sequencing. We therefore focus the remainder of this review on how newer sequencing approaches may enable identification of SVs in the clinic.

Prime suspects for clinical translation

Figure 1 provides an overview of common cancer related structural variation and the technologies currently available for detection. For detailed descriptions on structural variant classification and aetiology we direct the reader to recent reviews [24–27]. In the following section we focus on cancer promoting SVs that provide strong exemplars for clinical decision making.

Oncogenic fusions - Figure 1a

The prototypical oncogenic fusion, BCR-ABL in chronic myeloid leukaemia is the result of a reciprocal translocation between chromosomes 22 and 9 causing a BCR-ABL fusion protein[28]. Patients with this fusion respond to the tyrosine kinase inhibitor imatinib, which represents the earliest example of a rationally targeted cancer therapy[29]. Subsequent studies have uncovered specific fusions across a range of tumour types[30,31]. These fusions commonly arise either via translocation (e.g. BCR-ABL[28]), deletion (e.g. TMPRSS2-ERG[32]) or inversion (e.g. EML4-ALK[33]). In cases such as TMPRSS2 and ALK, the single gene may also have multiple possible fusion partners[32,34]. ALK fusions in non-small cell lung cancer are now a critical target for therapy. In contrast, intensive study on the TMPRSS2-ERG fusions which are found in up to 50% of prostate cancers[35], has not yet demonstrated therapeutic or functional insights. At present the majority of identified fusions in solid tumours have unknown function. These functional challenges mirror similar problems in identifying driver SNV mutations but it is important to emphasize that some cancers have specific mutator phenotypes that are selective for oncogenic fusions. For example, a subset of lung adenocarcinomas have been shown to be exclusively driven by fusions[36]. Conversely, fusions may also have classical loss of function effects and a recent study in prostate cancer detected an inversion causing a fusion of MSH2 and NRXN1, which inactivated MSH2 causing mis-match repair deficiency[37].

Clinical assays to identify fusions have focused on detection of known markers. Break-apart FISH, where probes located at either side of the breakpoint show a separation in the presence of the fusion, has been the assay of choice for clinical profiling. Where fusion specific antibodies exist, IHC has been applied as a simple low-cost alternative. Alternative approaches using sequencing based methods will depend highly on the recurrent nature of the underlying genomic aberration and the intron size of the two fused genes. If the aberration involves a loss of genetic material, then the detection task is easier as the absence of exons in the fusion genes is usually a good proxy for the presence of the fusion. If the aberration causing the fusion is balanced, detection becomes much more difficult[38]. If there

are breakpoint `hotspots' across tumours (i.e. the double-stranded breaks occur within kilobases of each other) then PCR primers can be designed either side of the break to detect the fusion using fresh or fixed DNA[39]. Otherwise, if the breaks occur in intermediate size regions, exon/intron capture[5] or long-range PCR[40] is required, providing that the DNA is of sufficient quality to yield long enough fragments[41]. Finally, for breaks which are not recurrent, paired-end WGS is the preferred approach.

It has, however, been shown that fusions can be detected using targeted sequencing either using a specific protocol which ligates nuclear proximal sequences for a gene of interest - TLA[42] or bioinformatically from chimeric reads in existing targeted sequencing data - BreaKmer[43]. It is also possible to interrogate RNA rather than DNA to detect fusions - if the breaks across patients result in the fusion of the same exons, targeted PCR based assays can be designed to pick up the fusion product in RNA, even from short fragments, with the caveat that the fusion must be expressed in the cells[44]. If only one of the fusion partners is known, PCR baits for the exons of the known partner can be paired with random hexamer priming to amplify the fragment for any fusion partner[45].

Oncogene amplification - Figure 1b

High-level gene amplification typically arises via focal copy-number change[46]. A canonical example in the clinic is the amplification of *ERBB2* in breast cancer patients, resulting in HER2 overexpression which can be treated with the antagonist trastuzumab[47]. Recently, *ERBB2* amplification has been demonstrated to occur in many tumour types, albeit at lower frequency, supporting the potential rational use of HER2 antagonist therapy[48]. Although, a recent clinical trial of off-label therapy use based on molecular characterisation showed no widespread benefit suggesting this strategy may be specific to certain therapies[49]. HER2 status is commonly assessed using IHC, with *ERBB2* amplification interrogated using FISH or CISH if IHC is uncertain[50]. Recently, it has been shown that sequencing based technologies such as ddPCR show greater clinical efficacy[51]. In many cases however, the interrogation of multiple genes simultaneously may improve clinical decision making (Box 1).

Targeted gene panels, exome and WGS based approaches all allow detection of amplifications across many chromosomal positions simultaneously, with varying accuracy and sensitivity. Methods for detection generally rely on read depth, heterozygous germline SNP allele frequencies, or a combination of both[52]. However, determining the precise number of copies can be difficult, especially for samples with low tumour cell content[5]. In these cases, a copy-number aberration will not be detected using low coverage sequencing unless the amplification is in excess of 6 copies[5]. Conversely, while targeted gene panels generally provide a read depth necessary for copy-number detection, artefacts such as PCR or hybridisation efficiency can skew the estimated copy-number. Despite these caveats, technical and algorithmic developments are yielding promising results, for example, the ability to sensitively detect *AR* and *CYP17A1* gains in the serum of castrate resistant prostate cancer patients[53].

Enhancer hijacking - Figure 1c

While there are not yet any examples of enhancer hijacking being targeted in the clinic, this mechanism provides a prototypical example for both the potential of large-scale tumour genomics studies to generate new targeted therapy avenues, and the associated complexities with functional interpretation. A recent study of whole-genome sequencing data from 137 medulloblastoma samples revealed clusters of SV breakpoints at 9q34 which correlated with strong upregulation of *GFI1B* in a subset of tumours[54]. Further investigation showed these SVs, made up of various classes, all juxtaposed active enhancers to GFI1B, increasing expression of the proto-oncogene. While this work reveals a potential therapeutic target for treatment of a subset of medulloblastomas, routine detection of these cases would currently require whole-genome sequencing coupled with gene expression profiling - a costly and labour intensive diagnostic. Until WGS is widely adopted, this class of SV is likely to remain undetected in a clinical setting.

Tumour suppressor deletion - Figure 1d

Deletions are the most complex class of SVs to detect and interpretation is two-fold as confirmed inactivation of both copies is required. There are four common paths to

tumour suppressor inactivation: 1) a deleterious heterozygous germline mutation coupled with loss of the functional copy, such as BRCA1 in breast cancer[55]; 2) a somatic point mutation in one allele, followed by loss of the remaining copy, such as TP53 in ovarian cancer[14]; 3) epigenetic silencing of one copy followed by loss of the functional copy, such as *MLH1* in colorectal cancer[56]; or 4) homozygous loss of both copies via two separate deletion events.

Deletions of protein coding genes causing loss of function are typically diagnosed using IHC. However, IHC is commonly not applicable owing to lack of validated clinical antibodies (e.g. detection of NF1 protein) or when the functional impact is a result of a non-coding change such as the loss of a microRNA with unknown protein target[57]. To confirm a loss of function with sequencing, two assays may be required, one to confirm the copy-number change and another to confirm the secondary hit on the remaining copy[58]. Depending on the distribution of heterozygous germline SNPs in the region of interest, determining LOH in tissue samples may require an estimate of the tumour purity (fraction of tumour cells relative to normal) and ploidy. This can usually only be estimated using genome-wide profiling techniques (array or sequencing based) in combination with computational algorithms such as ASCAT[59] or ABSOLUTE[60] which can determine the most likely purity and ploidy state. Given these limitations, the use of sequencing-based assays for detecting deletions will not be sufficient for unambiguous detection of loss of function without further advances in technology and clinical workflow.

Genomic instability - Figure 1e

Integrating data from multiple SVs using genome-wide profiling can reveal diagnostic and prognostic information which now offer clinically useful biomarkers [61–63] in contrast to the current uncertainty about which single SV event is dominant or actionable (discussed below). Scores for focal complex rearrangements such as the complex arm aberration index (CAAI) have prognostic power in breast cancer[64] and have been shown to be robust across larger cohorts of breast and high grade serous ovarian cancer cases[65]. In gastrointestinal tumours, a Genomic Index (GI) integrating the number and type of copy-number changes has been linked with

prognosis[66], and may be useful for molecularly stratifying intermediate risk disease [67]. In addition other integrative measures of genomic instability developed from loss of heterozygosity (LOH), telomeric allelic imbalance (TAI), and large-scale state transitions (LST) have been correlated with underlying homologous recombination deficiency in breast cancer and related sensitivity to platinum-based chemotherapy [68] and subsequently commercialized. Similar approaches are also in development for predicting response to PARP inhibitors in high grade serous ovarian cancer [69,70]. Further validation of these signatures in large cohorts is now required together with comparison to other signatures that may reflect specific aberrant DNA damage and repair, similar to the SNV signatures identified using non-matrix factorization[71].

While these approaches use complex rearrangements to compute global measures of instability, it is not yet possible to accurately resolve complex events to the point where underlying drivers can be determined. However, by integrating short-read sequencing data with 10X Genomics synthetic long-read technology and BioNano Genomics mapping, it may be feasible to resolve these events from the resulting phased, long-reads[72].

Technology frontrunners

Balancing cost and sensitivity: shallow whole-genome sequencing

Advances in methods that adjust for GC sequencing bias, along with careful curation ("black-listing") of problematic genomic regions has dramatically improved depth of coverage approaches for genome-wide copy number profiling[73]. It is now possible to obtain a meaningful genome-wide copy-number profile from sequencing depths as low as 0.1×, applied to FFPE or frozen samples[73]. If matched germline samples are processed, it is possible to discern germline from somatic copy-number alterations. However, even in the absence of matched germline it is still possible to extract reliable copy-number profiles with clinical utility (see Box 4). In addition, if the DNA is of sufficient quality, long-insert paired-end shallow whole-genome sequencing can be used to obtain both copy-number and balanced rearrangement information[74]. It is has also been shown that these shallow whole-genome sequencing (sWGS) techniques can be used to interrogate circulating tumour DNA in

plasma[75]. These advances pave the way for a cost effective genome-wide profiling strategy for monitoring copy-number changes in tumours in response to therapy.

One caveat of these techniques in the cancer setting is that their performance is tightly coupled to both tumour purity and ploidy of the sample[76]. Furthermore, a critical feature for monitoring therapy response is the ability to observe and account for tumour heterogeneity[77]. In the case of copy-number profiling, tumour heterogeneity can be observed as a mixture of copy-number states which result in a non-integer copy-number. However, to decide whether a copy-number is non-integer, absolute copy-number changes rather than relative copy-number changes must be observed. For this, estimates of purity and ploidy are required[59]. In the absence of deeper WGS, reasonable estimates for these can be obtained via histopath and FISH/CISH, or bioinformatically inferred[60]. If available, we highlight in Box 2 that subclonal copy number changes are theoretically detectable with a sWGS strategy. One cost-effective approach is to couple sWGS with exome seq facilitating absolute copy-number calling[73].

Getting more bang for the buck: advanced bioinformatics

Advances in bioinformatic approaches have boosted our ability to extract useful information out of relatively low-coverage sequencing experiments. For example, exome and targeted gene panels that were originally designed to interrogate only SNVs and INDELs, are now being extended to estimates of copy-number[78,79]. Furthermore, a perceived defect of exome capture, namely inefficient hybridization causing “off target” reads, has been transformed into a useful data source for reconstructing genome-wide copy-number number profiles[80].

Other improved algorithms also make it possible to go beyond the simple binary score indicating presence or absence of a given mutation—by using the allele frequency of a mutation it is possible reconstruct the evolution of a tumour and determine the fraction of tumour cells that contain the mutation, even for SVs[37,81]. This may have strong clinical benefits, as detecting actionable mutations that are present on the trunk of the tumour phylogenetic tree rather than the branches, may allow selection of therapy that targets the bulk of the tumour cells[77].

Some pitfalls

Beyond the technical issues associated with detection, there are other factors which obscure clinical decision making using SVs:

More than one driver mutation

In Box 1 we use a hypothetical scenario to illustrate the challenges faced when a copy-number change is observed across multiple drivers, of which only one is the true driver. In this case, neither WGS or targeted sequencing can resolve the dilemma, each with their own shortcomings. The situation becomes worse when dealing with unstable genomes, where complex rearrangements can result in a list of putative drivers ranging in the 1000s. The SNV profiling field has faced a similar challenge, thus many of the methods developed can be adapted to SVs[82]. Approaches that use gene and protein interaction networks to elucidate key driver pathways hit by multiple mutations, offer a promising technique for narrowing the list of putative driver events to the point where a targeted therapy can be rationalised (methods reviewed in detail in [83]). Alternatively, integrative SV analysis (see text above) may offer treatment choices for some patients.

To assist in the process of therapy rationalisation, many cancer centres have appointed panels of experts to decide on the best course of treatment given complex molecular results[84]. These panels are typically made up of clinicians, scientists, bioinformaticians and others that collectively decide if the molecular evidence is sufficient to make a therapeutic intervention. While this pipeline results in a high-level of care, it is ultimately low-throughput. One way to overcome this is to ensure that detailed information on the decision making process of these panels is captured so that areas of redundancy and automation can be identified and throughput improved. One of the critical areas for achieving this improvement is enhanced annotation of structural variants and the functional impact (including possible confounders). To help address this,[85] has proposed the introduction of a clinical targetability index, which is supported by databases of manually curated druggable mutations including the incorporation of initiatives such as My Cancer Genome[https://www.mycancergenome.org], Targeted Cancer

Care[https://targetedcancercare.massgeneral.org], and Personalized Cancer Therapy[https://pct.mdanderson.org]. Measures such as these will assist clinicians in deciding which aberrations to target.

Tumour heterogeneity

Tumour heterogeneity is another factor that impacts strongly on the clinical interpretation of mutations. Discussed extensively recently[83,86] tumour heterogeneity has a profound effect on the choice of sequencing approach for interrogating SVs in tumours. While WGS provides a comprehensive view of the genomic makeup of a tumour, this is limited in absolute depth, revealing only mutations found in the bulk of tumour cells. In contrast, targeted sequencing provides a deeper, narrow view, with the potential benefit of detecting tumour heterogeneity for single mutations. This may be critical in cancers where identification of initially, small, resistant populations of cells is paramount. A recent review by Hiley et al. [77] highlights how modelling tumour heterogeneity and understanding the life history of a tumour can assist with prioritizing therapeutic targets. By targeting mutations that arise early in tumour development present in all cells, the bulk of the tumour will be hit by the treatment. Alternatively, by characterising cells with different driver mutations, combination therapies can be designed so that all observed drivers are targeted. Methods for identifying heterogeneous copy-number aberrations[87] and SVs[37], are likely to have a significant impact on improved decision making (See Box 3 for an example of how clonality analysis is currently being used to inform therapy in the clinic). Currently, though, the majority of these methods are designed to use deep WGS as input and further technical and algorithmic developments are required for routine assessment of copy-number heterogeneity using targeted or shallow sequencing strategies.

Necessary hurdles

A putative sequencing based biomarker must be subjected to rigorous testing of its analytical validity, clinical validity and clinical utility[88] before widespread adoption in the clinic. In the context of SVs, this would typically mean proving that the SV can be accurately and robustly detected (analytical validity); showing that the SV associates with the clinical outcome of interest, in this case, target gene activity in the tumour

(clinical validity); and demonstrating that the detection of the SV leads to a targeted therapy which improves patient outcomes (clinical utility). These are major challenges with even analytical validation presenting a significant technical and analytical challenge. For the validation of the copy number calling for the FoundationOne assay, multiple pools of mixed normal and tumour cell line DNA were used in ratios of 20-75% tumour content for blinded calling of focal gene amplification and homozygous deletion across repeat experiments[5]. Establishing clinical validity for a novel SV will continue to be an expensive and time consuming process as it relies upon large validation sample sets with orthogonal genomic characterization with RNA profiling and protein assays. One approach to help mitigate this may be the use of careful sequencing study designs which are well powered to inform clinical decision making[2]. Furthermore, careful certification and accreditation of the bioinformatics pipelines for processing sequencing data is required to ensure analytical validity[89]. Tools such as Docker, which completely encapsulate all software needed for the analysis in a virtual machine (https://www.docker.com/) are improving reproducibility, testing and deployment. SV specific accreditation of pipelines can draw on the lessons learnt from certification of SNV pipelines[90], as well as guidelines developed for germline testing[91]. Promising examples of approaches for somatic detection are emerging such as the system used by Princess Margaret Cancer Centre for clinical somatic variant classification[92], however, it is important that SV profiling be integrated in these systems early in their development as many of the underlying quality control principles of sequencing use in the clinic apply in both cases.

To WGS or not to WGS

Since the goal of sequencing a complete genome for $1000 was realised in 2014 [93], there has been continuing debate on whether the time is right to adopt WGS routinely in the clinic[94]. For reasons outlined above, comprehensive characterisation of SVs in specific cancers would greatly benefit from routine deep WGS, however, significant barriers still need to be overcome before this could be considered feasible. These include a reduction in the high human costs associated with computational analysis, functional interpretation, and identification of actionable

drivers. Developing strategies to overcome these challenges is currently uninformed as there are no studies that have directly assessed the clinical benefit of whole-genome sequencing in cancer although many major cancer centres are starting to grapple with the significant infrastructure required for clinical WGS. In addition, several national and regional WGS sequencing efforts are underway that may mitigate the analysis bottlenecks by economies of scale, albeit by imposition of pragmatic or restricted bioinformatic reporting to achieve clinically useful turn around times. In the UK the Genomics England 100,000 Genomes Project has now sequenced 11,221 genomes from NHS patients being investigated for either rare inherited disease or somatic sequencing at cancer diagnosis. Similar projects are ongoing in California, Vancouver and the Netherlands.

A key technical challenge to overcome in the cancer setting is to be able to process FFPE material. As far as we are aware, there is no deep WGS published with DNA isolated from FFPE clinical material. In our hands the sequence yield has been too low from a single sequence lane of the Illumina X10 to allow genome-wide mutation analysis. However, the yield was sufficient to perform genome-wide mapping of SVs (translocation and copy numbers). These challenges can be mitigated by different bioinformatic methods, however, it is important to acknowledge the overhead required for the development of bioinformatic approaches which provide clinical grade mutation calling. In this regard it is particularly instructive to review lessons learnt from developing SNV calling algorithms. Initial “state-of-the-art” calling methods developed in academia showed significant discrepancies[95], and it is only recently we have been able to accurately quantify the performance of the different approaches via international mutation calling challenges[96]. A significant impact of the Pan Cancer Analysis of Whole Genomes project (https://dcc.icgc.org/pcawg) has been the curation of consensus calling strategies using deep WGS from over 2800 tumours, and with this resource we are now in a position to develop and validate calls that can be deemed robust enough for clinical work.

While we still have a long way to go before sufficient capability is acquired for widespread deployment of WGS in the clinic, there are select scenarios where WGS is already being adopted. Box 3 and Box 4 illustrate two cases where a sWGS

strategy is replacing aCGH and FISH for assessing copy-number in the clinic. Another likely candidate for early adoption of WGS in the clinic is genome-wide characterisation of loss-of-heterozygosity. Knowledge of this, combined with identification of SNV drivers assists in robust interpretation of putative targets.

Concluding remarks

Although we expect extensive, ongoing debate on the role of DNA sequencing in the clinic[97], we see a clear need for cheap and accessible sequencing-based approaches in the clinic to interrogate SVs and to widen rational therapeutic choices. These approaches will provide a stronger basis for understanding the genomic architecture of cancers and should be integrated into large-scale discovery efforts for clinical biomarkers. In the short term we expect these tools will be used for orthogonal validation for validated relevant SVs, however over the long term, we expect modifications of sWGS to become the primary tool for SV detection. Although deep whole-genome sequencing will provide the most comprehensive approach, care needs to be taken to develop the correct infrastructure to ensure both test and analysis costs remain low[98] and practitioners are given sufficient education in interpretation of sequencing based tests[99]. Until then, a mixed strategy of shallow WGS and targeted sequencing is likely to be a sufficient framework for diagnosis, prognosis and tracking treatment resistance.

Acknowledgements

This work was supported by Cancer Research UK [grant numbers A15973, A15601: G.M, J.D.B], VUmc Cancer Center Amsterdam [VUmc-CCA: BY] and the Dutch Cancer Society [VU 2015-7882: BY]. We would like to thank Daoud Sie for help in preparing Figure ii.

References

1 Garraway, L.A. and Lander, E.S. (2013) Lessons from the cancer genome. *Cell* 153, 17–37

2 Mwenifumbo, J.C. and Marra, M.A. (2013) Cancer genome-sequencing study design. *Nat. Rev. Genet.* 14, 321–32

3 Marx, V. (2014) Cancer genomes: discerning drivers from passengers. *Nat. Methods* 11, 375–379

4 Foley, S.B. *et al.* (2015) Use of Whole Genome Sequencing for Diagnosis and Discovery in the Cancer Genetics Clinic. *EBioMedicine* 2, 74–81

5 Frampton, G.M. *et al.* (2013) Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat. Biotechnol.* 31, 1023–31

6 Yan, H. *et al.* (2009) IDH1 and IDH2 mutations in gliomas. *N. Engl. J. Med.* 360, 765–73

7 Ceccarelli, M. *et al.* (2016) Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell* 164, 550–563

8 Cheng, D.T. *et al.* (2015) MSK-IMPACT: A Hybridization Capture-Based Next-Generation Sequencing Clinical Assay for Solid Tumor Molecular Oncology. *J. Mol. Diagn.* 17, 251–264

9 Johnson, D.B. *et al.* (2014) Enabling a genetically informed approach to cancer medicine: a retrospective evaluation of the impact of comprehensive tumor profiling using a targeted next-generation sequencing panel. *Oncologist* 19, 616–22

10 Ramkissoon, S.H. *et al.* (2015) Clinical implementation of integrated whole-genome copy number and mutation profiling for glioblastoma. *Neuro. Oncol.* DOI: 10.1093/neuonc/nov015

11 Louis, D.N. *et al.* (2014) International Society Of Neuropathology--Haarlem consensus guidelines for nervous system tumor classification and grading. *Brain Pathol.* 24, 429–35

12 Rehm, H.L. (2013) Disease-targeted sequencing: a cornerstone in the clinic. *Nat. Rev. Genet.* 14, 295–300

13 Ciriello, G. *et al.* (2013) Emerging landscape of oncogenic signatures across human cancers. *Nat. Genet.* 45, 1127–1133

14 Bell, D. *et al.* (2011) Integrated genomic analyses of ovarian carcinoma. *Nature* 474, 609–615

15 Cheng, C. *et al.* (2016) Whole-Genome Sequencing Reveals Diverse Models of Structural Variations in Esophageal Squamous Cell Carcinoma. *Am. J. Hum. Genet.* 98, 256–274

16 Pugh, T.J. *et al.* (2013) The genetic landscape of high-risk neuroblastoma. *Nat. Genet.* 45, 279–284

17 George, J. *et al.* (2015) Comprehensive genomic profiles of small cell lung cancer. *Nature* 524, 47–53

18 Koboldt, D.C. *et al.* (2012) Comprehensive molecular portraits of human breast tumours. *Nature* 490, 61–70

19 Bang, Y.-J. *et al.* (2010) Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. *Lancet (London, England)* 376, 687–97

20 Mertens, F. *et al.* Cytogenetic analysis in the examination of solid tumors in children. *Pediatr. Hematol. Oncol.* 11, 361–77
21 Hastings, R.J. *et al.* (2016) Guidelines for cytogenetic investigations in tumours. *Eur. J. Hum. Genet.* DOI: 10.1038/ejhg.2015.35
22 Editorial (2013) Consortium launches genotyping effort. *Cancer Discov.* 3, 1321–2
23 Commo, F. *et al.* (2015) Impact of centralization on aCGH-based genomic profiles for precision medicine in oncology. *Ann. Oncol.* 26, 582–8
24 Alkan, C. *et al.* (2011) Genome structural variation discovery and genotyping. *Nat. Rev. Genet.* 12, 363–76
25 Escaramis, G. *et al.* (2015) A decade of structural variants: description, history and methods to detect structural variation. *Brief. Funct. Genomics* DOI: 10.1093/bfgp/elv014
26 Tubio, J.M.C. (2015) Somatic structural variation and cancer. *Brief. Funct. Genomics* DOI: 10.1093/bfgp/elv016
27 Tattini, L. *et al.* (2015) Detection of Genomic Structural Variants from Next-Generation Sequencing Data. *Front. Bioeng. Biotechnol.* 3, 92
28 ROWLEY, J.D. (1973) A New Consistent Chromosomal Abnormality in Chronic Myelogenous Leukaemia identified by Quinacrine Fluorescence and Giemsa Staining. *Nature* 243, 290–293
29 Druker, B.J. *et al.* (2001) Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N. Engl. J. Med.* 344, 1031–7
30 Davare, M.A. and Tognon, C.E. (2015) Detecting and targeting oncogenic fusion proteins in the genomic era. *Biol. Cell* DOI: 10.1111/boc.201400096
31 Edwards, P.A.W. (2010) Fusion genes and chromosome translocations in the common epithelial cancers. *J. Pathol.* 220, 244–54
32 Tomlins, S.A. *et al.* (2005) Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* 310, 644–8
33 Soda, M. *et al.* (2007) Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* 448, 561–6
34 Shaw, A.T. *et al.* (2013) Tyrosine kinase gene rearrangements in epithelial malignancies. *Nat. Rev. Cancer* 13, 772–87
35 Tomlins, S.A. *et al.* (2008) Role of the TMPRSS2-ERG gene fusion in prostate cancer. *Neoplasia* 10, 177–88
36 Saito, M. *et al.* (2015) Development of Lung Adenocarcinomas with Exclusive Dependence on Oncogene Fusions. *Cancer Res.* DOI: 10.1158/0008-5472.CAN-14-3282
37 Hong, M.K.H. *et al.* (2015) Tracking the origins and drivers of subclonal metastatic expansion in prostate cancer. *Nat. Commun.* 6, 6605
38 Talkowski, M.E. *et al.* (2011) Next-generation sequencing strategies enable routine detection of balanced chromosome rearrangements for clinical diagnostics and genetic research. *Am. J. Hum. Genet.* 88, 469–81
39 van Dongen, J.J.M. *et al.* (2003) Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia* 17, 2257–317
40 Jia, H. *et al.* (2014) Long-range PCR in next-generation sequencing: comparison of six enzymes and evaluation on the MiSeq sequencer. *Sci. Rep.*

4, 5737
41 Blow, N. (2007) Tissue preparation: Tissue issues. *Nature* 448, 959–63
42 de Vree, P.J.P. *et al.* (2014) Targeted sequencing by proximity ligation for comprehensive variant detection and local haplotyping. *Nat. Biotechnol.* 32, 1019–25
43 Abo, R.P. *et al.* (2015) BreaKmer: detection of structural variation in targeted massively parallel sequencing data using kmers. *Nucleic Acids Res.* 43, e19
44 Lee, M.S. *et al.* (1989) Detection of two alternative bcr/abl mRNA junctions and minimal residual disease in Philadelphia chromosome positive chronic myelogenous leukemia by polymerase chain reaction. *Blood* 73, 2165–70
45 Zheng, Z. *et al.* (2014) Anchored multiplex PCR for targeted next-generation sequencing. *Nat. Med.* 20, 1479–84
46 Albertson, D.G. (2006) Gene amplification in cancer. *Trends Genet.* 22, 447–55
47 Slamon, D.J. *et al.* (2001) Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N. Engl. J. Med.* 344, 783–92
48 Yan, M. *et al.* (2015) HER2 expression status in diverse cancers: review of results from 37,992 patients. *Cancer Metastasis Rev.* 34, 157–64
49 Le Tourneau, C. *et al.* (2015) Molecularly targeted therapy based on tumour molecular profiling versus conventional therapy for advanced cancer (SHIVA): a multicentre, open-label, proof-of-concept, randomised, controlled phase 2 trial. *Lancet. Oncol.* 16, 1324–34
50 Rakha, E.A. *et al.* (2015) Updated UK Recommendations for HER2 assessment in breast cancer. *J. Clin. Pathol.* 68, 93–9
51 Kinugasa, H. *et al.* (2015) Droplet digital PCR measurement of HER2 in patients with gastric cancer. *Br. J. Cancer* DOI: 10.1038/bjc.2015.129
52 Liu, B. *et al.* (2013) Computational methods for detecting copy number variations in cancer genome using next generation sequencing: principles and challenges. *Oncotarget* 4, 1868–81
53 Salvi, S. *et al.* (2015) Circulating cell-free AR and CYP17A1 copy number variations may associate with outcome of metastatic castration-resistant prostate cancer patients treated with abiraterone. *Br. J. Cancer* DOI: 10.1038/bjc.2015.128
54 Northcott, P.A. *et al.* (2014) Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* 511, 428–34
55 Merajver, S.D. *et al.* (1995) Germline BRCA1 mutations and loss of the wild-type allele in tumors from families with early onset breast and ovarian cancer. *Clin. Cancer Res.* 1, 539–44
56 Herman, J.G. *et al.* (1998) Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc. Natl. Acad. Sci. U. S. A.* 95, 6870–5
57 Varambally, S. *et al.* (2008) Genomic loss of microRNA-101 leads to overexpression of histone methyltransferase EZH2 in cancer. *Science* 322, 1695–9
58 Leary, R.J. *et al.* (2008) Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers. *Proc. Natl. Acad. Sci. U. S. A.* 105, 16224–9

59 Van Loo, P. *et al.* (2010) Allele-specific copy number analysis of tumors. *Proc. Natl. Acad. Sci. U. S. A.* 107, 16910–5
60 Carter, S.L. *et al.* (2012) Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* 30, 413–21
61 Carén, H. *et al.* (2010) High-risk neuroblastoma tumors with 11q-deletion display a poor prognostic, chromosome instability phenotype with later onset. *Proc. Natl. Acad. Sci. U. S. A.* 107, 4323–8
62 Fridlyand, J. *et al.* (2006) Breast tumor copy number aberration phenotypes and genomic instability. *BMC Cancer* 6, 96
63 Chin, S.F. *et al.* (2007) High-resolution aCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. *Genome Biol.* 8, R215
64 Russnes, H.G. *et al.* (2010) Genomic architecture characterizes tumor progression paths and fate in breast cancer patients. *Sci. Transl. Med.* 2, 38ra47
65 Vollan, H.K.M. *et al.* (2015) A tumor DNA complex aberration index is an independent predictor of survival in breast and ovarian cancer. *Mol. Oncol.* 9, 115–27
66 Lagarde, P. *et al.* (2012) Mitotic checkpoints and chromosome instability are strong predictors of clinical outcome in gastrointestinal stromal tumors. *Clin. Cancer Res.* 18, 826–38
67 Lartigue, L. *et al.* (2015) Genomic index predicts clinical outcome of intermediate-risk gastrointestinal stromal tumours, providing a new inclusion criterion for imatinib adjuvant therapy. *Eur. J. Cancer* 51, 75–83
68 Telli, M.L. *et al.* (2016) Homologous Recombination Deficiency (HRD) Score Predicts Response to Platinum-Containing Neoadjuvant Chemotherapy in Patients with Triple-Negative Breast Cancer. *Clin. Cancer Res.* DOI: 10.1158/1078-0432.CCR-15-2477
69 Abkevich, V. *et al.* (2012) Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. *Br. J. Cancer* 107, 1776–82
70 McNeish, I.A. *et al.* (2015) Results of ARIEL2: A Phase 2 trial to prospectively identify ovarian cancer patients likely to respond to rucaparib using tumor genetic analysis. *J. Clin. Oncol.* 33,
71 Alexandrov, L.B. *et al.* (2013) Signatures of mutational processes in human cancer. *Nature* 500, 415–21
72 Mostovoy, Y. *et al.* (2016) A hybrid approach for de novo human genome sequence assembly and phasing. *Nat. Methods* 13, 587–590
73 Scheinin, I. *et al.* (2014) DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Res.* 24, 2022–2032
74 Liang, W.S. *et al.* (2014) Long insert whole genome sequencing for copy number variant and translocation detection. *Nucleic Acids Res.* 42, e8
75 Heitzer, E. *et al.* (2013) Tumor-associated copy number changes in the circulation of patients with prostate cancer identified through whole-genome sequencing. *Genome Med.* 5, 30
76 Cibulskis, K. *et al.* (2013) Sensitive detection of somatic point mutations in

impure and heterogeneous cancer samples. *Nat. Biotechnol.* 31, 213–9
77 Hiley, C. *et al.* (2014) Deciphering intratumor heterogeneity and temporal acquisition of driver events to refine precision medicine. *Genome Biol.* 15, 453
78 Wang, C. *et al.* (2014) PatternCNV: a versatile tool for detecting copy number changes from exome sequencing data. *Bioinformatics* 30, 2678–80
79 Boeva, V. *et al.* (2014) Multi-factor data normalization enables the detection of copy number aberrations in amplicon sequencing data. *Bioinformatics* 30, 3443–50
80 Kuilman, T. *et al.* (2015) CopywriteR: DNA copy number detection from off-target sequence data. *Genome Biol.* 16, 49
81 Ha, G. *et al.* (2014) TITAN: inference of copy number architectures in clonal cell populations from tumor whole-genome sequence data. *Genome Res.* 24, 1881–93
82 Quintáns, B. *et al.* (2014) Medical genomics: The intricate path from genetic variant identification to clinical interpretation. *Appl. Transl. Genomics* 3, 60–67
83 Raphael, B.J. *et al.* (2014) Identifying driver mutations in sequenced cancer genomes: computational approaches to enable precision medicine. *Genome Med.* 6, 5
84 Simon, R. and Roychowdhury, S. (2013) Implementing personalized cancer genomics in clinical trials. *Nat. Rev. Drug Discov.* 12, 358–69
85 Dienstmann, R. *et al.* (2015) Database of Genomic Biomarkers for Cancer Drugs and Clinical Targetability in Solid Tumors. *Cancer Discov.* 5, 118–123
86 McGranahan, N. and Swanton, C. (2015) Biological and Therapeutic Impact of Intratumor Heterogeneity in Cancer Evolution. *Cancer Cell* 27, 15–26
87 Oesper, L. *et al.* (2013) THetA: inferring intra-tumor heterogeneity from high-throughput DNA sequencing data. *Genome Biol.* 14, R80
88 Parkinson, D.R. *et al.* (2014) Evidence of clinical utility: an unmet need in molecular diagnostics for patients with cancer. *Clin. Cancer Res.* 20, 1428–44
89 Sboner, A. and Elemento, O. (2015) A primer on precision medicine informatics. *Brief. Bioinform.* DOI: 10.1093/bib/bbv032
90 Oliver, G.R. *et al.* (2015) Bioinformatics for clinical next generation sequencing. *Clin. Chem.* 61, 124–35
91 Gargis, A.S. *et al.* (2012) Assuring the quality of next-generation sequencing in clinical laboratory practice. *Nat. Biotechnol.* 30, 1033–6
92 Sukhai, M.A. *et al.* (2015) A classification system for clinical relevance of somatic variants identified in molecular profiling of cancer. *Genet. Med.* DOI: 10.1038/gim.2015.47
93 Check Hayden, E. (2014) Is the $1,000 genome for real? *Nature* DOI: 10.1038/nature.2014.14530
94 Check Hayden, E. (2014) Genome sequencing stumbles towards the clinic. *Nature* DOI: 10.1038/nature.2014.14842
95 Kim, S. *et al.* (2013) Comparing somatic mutation-callers: beyond Venn diagrams. *BMC Bioinformatics* 14, 189
96 Ewing, A.D. *et al.* (2015) Combining tumor genome simulation with crowdsourcing to benchmark somatic single-nucleotide-variant detection. *Nat. Methods* 12, 623–630
97 Tripathy, D. *et al.* (2014) Next generation sequencing and tumor mutation profiling: are we ready for routine use in the oncology clinic? *BMC Med.* 12,

140
98 Dewey, F.E. *et al.* (2014) Clinical interpretation and implications of whole-genome sequencing. *JAMA* 311, 1035–45
99 Yan, B. *et al.* (2014) Integrating translational bioinformatics into the medical curriculum. *Int. J. Med. Educ.* 5, 132–4
100 Krijgsman, O. *et al.* (2013) Detection limits of DNA copy number alterations in heterogeneous cell populations. *Cell. Oncol. (Dordr).* 36, 27–36
101 Ostrovnaya, I. *et al.* (2011) Clonality: an R package for testing clonal relatedness of two tumors from the same patient based on their genomic profiles. *Bioinformatics* 27, 1698–9
102 Gallegos Ruiz, M.I. *et al.* (2007) Genetic heterogeneity in patients with multiple neoplastic lung lesions: a report of three cases. *J. Thorac. Oncol.* 2, 12–21
103 Chiu, R.W.K. *et al.* (2008) Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. *Proc. Natl. Acad. Sci.* 105, 20458–20463
104 Rajmohan, K.S. *et al.* (2016) Prognostic significance of histomolecular subgroups of adult anaplastic (WHO Grade III) gliomas: applying the “integrated” diagnosis approach. *J. Clin. Pathol.* DOI: 10.1136/jclinpath-2015-203456
105 Katsanis, S.H. and Katsanis, N. (2013) Molecular genetic testing and the future of clinical genomics. *Nat. Rev. Genet.* 14, 415–26

Text Boxes

Box 1 - Oncogene amplification: a clinical dilemma (hypothetical)

Consider a hypothetical situation where a patient presenting with non-small cell lung cancer has their tumour sequenced using a targeted capture panel for all exons of EGFR. The results come back negative for SNVs and INDELs, however, the assay reports a copy-number gain spanning EGFR which can be targeted with a tyrosine kinase inhibitor[29]. The obvious temptation here is to commence treatment with a tyrosine kinase inhibitor such as erlotinib. However, what we don't know in this case, is that the EGFR gain is a result of a whole chromosome level change, which happens to incorporate the true tumour driver, MET, also on chromosome 7. MET amplification is known to confer resistance to erlotinib, but can be targeted with crizotinib. In this case, only a more comprehensive assay would have allowed us to identify the true target, although this test would have come at an increased cost. Furthermore, even if this assay was carried out, we would be faced with another dilemma: of the two putative targets, which is the true driver gene? Which is likely to be the most effective treatment? For the capture panel results, there is only one possible treatment option, which in this case may have resulted in a failed response, negatively impacting on the perceived efficacy of the treatment and test. Whereas, in a whole-genome case, uncertainty around which is the true driver is likely to result in no treatment intervention unless additional testing is carried out to determine the true driver. Which is the best assay?

Box 2 - sWGS for monitoring subclonal copy-number changes

The ability to detect subclonal copy-number changes (changes present in a subset of tumour cells in a sample) is important for measuring intra-tumour heterogeneity and fluctuations in responsive and resistant clones during therapy. Here we show that, in theory, sWGS has the power to cheaply and robustly detect subclonal copy-number changes. By performing power calculations (https://gmacintyre.shinyapps.io/sWGS_power/), we plot the number of reads required to detect significant copy-number changes with 80% power as a function of tumour purity (Figure i). If we assume a modest tumour purity of 54% (the sample

contains 54% tumour cells, 46% normal) and an average ploidy of 2, we show that 20 million reads is sufficient to detect 1 megabase copy-number changes at subclonal fractions as low as 20%. This is inline with similar calculations carried out for arrays[100]. This example highlights the benefit of adopting an affordable sWGS strategy to measure copy-number changes linked to therapy response.

Box 3 - Clonality screening program using copy-number profiling at VUmc clinic, Amsterdam

The clonal relationship between tumours can be interpreted in an automated fashion by calculating a likelihood ratio to distinguish tumor pairs[101]. This is used to assist therapy management of patients with non-small cell lung cancer at the VUmc clinic in Amsterdam. Patients which present with multiple or secondary tumors, have their tumours profiled to determine whether they stem from the original primary tumor or are independent tumors[102]. If tumors are clonal and derived from the primary, the initial therapy is considered to have been ineffective and treatment is not repeated. Whereas if the tumor is novel, therapy is continued in the hope that this tumour will also respond. A low incidence rate (± 1 case/week at VUmc) combined with a requirement for a fast turnover time (7 days from DNA isolations to diagnosis) implies high workload, since single cases have to be processed separately. To reduce per sample cost, a transition from aCGH to sWGS is currently taking place, however, samples will still have to be run individually given the aforementioned low incidence rate and required turnover time. To overcome this we are currently using existing infrastructure that also performs sWGS for non-invasive prenatal testing (NIPT, [103]). At writing of this review we have finalized the validation phase for accreditation purposes to transition from arrays to sWGS for the clonality analysis, where the assay has been run in parallel on both platforms. This transition from arrays to sWGS will be an enormous time and cost saving measure, whilst adhering to the same high diagnostic standards required by the accreditation measures applied in the lab (CCKL accredited: ISO 15189:2007 certificate 111).

Box 4 - Diagnosis of low grade gliomas warrants genome-wide copy number profiling

In low-grade gliomas, 1p and 19q codeletion is currently recommended as a prognostic indicator for use in the clinic [11]. If present, a watchful waiting approach is adopted rather than immediate aggressive treatment. For this diagnosis, whole arm losses of 1p and 19q are mandatory; however common tests currently in use, use FISH probes which sample one or few chromosomal locations[21]. Importantly, it is the loss of whole chromosome arms, not parts, that are the true prognostic indicators. Moreover, it has been shown that inter-observer variability analysis of "oligo-dendroglioma" using histo-pathology (only) would be highly reduced by measuring the clonal 1p/19q co-deletion in these tumors [11,104]. These factors combined make a strong case for switching to a more comprehensive genome-wide copy-number profiling strategy[7]. At the time of writing, the VUmc in Amsterdam, that routinely carries out this diagnostic is opting to run all low-grade gliomas samples using both sWGS (see Figure ii for an example) in parallel with the current FISH test and side-by-side (thus same sequence lanes) with the NIPT and clonality tests outlined in Box 3.

Glossary

aCGH: array Comparative Genomic Hybridisation. A DNA microarray used for determine copy-number, typically genome-wide.

BreaKmer: A bioinformatic algorithm which uses chimeric reads in existing sequencing data to determine the existence of structural variant breakpoints.

CISH: Chromogenic In Situ Hybridization. A cytogenetic technique where probes labelled with biotin or digoxigenin are hybridised to specific regions of the genome and observed under a widefield microscope.

ddPCR: Digital droplet polymerase chain reaction. PCR performed on a single molecule that has been isolated in a water-oil emulsion droplet.

FFPE: Formalin fixed paraffin embedded. A term used to describe tissue samples that have been fixed in a formalin solution and embedded in paraffin wax blocks.

FISH: Fluorescent In Situ Hybridisation. A cytogenetic technique where fluorescently labelled probes are hybridised to specific regions of the genome and observed under a fluorescent microscope.

IHC: Immunohistochemistry. The detection of antigens from tissue sections using specific antibodies that bind the antigen with either chromogenic or fluorescent reporters.

INDEL: Insertion/deletion (somatic). A small stretch of base-pairs, typically 3-10 in length that is either deleted or inserted in the tumour genome.

MLPA: Multiplex ligation-dependent Probe Amplification. A multiplex PCR assay for detecting changes in copy-number.

PCR: Polymerase chain reaction. A molecular technique used to amplify DNA.

SNP array: Single nucleotide polymorphism array. A DNA microarray used to call germline variation that can also be used to determine copy-number.

SNV: Single-nucleotide variant (somatic). A mutation resulting in the change of a single base-pair in the genome of a tumour cell.

SV: Structural variant (somatic). A double stranded break in the DNA of a tumour cell resulting in either a balanced rearrangement where no DNA is lost such as an inversion or translocation, or in an unbalanced rearrangement such as a deletion or amplification (also known as a copy-number aberration).

sWGS: Shallow whole-genome sequencing. Short read sequencing of tumour genomes with limited read depth, typically less than 1x coverage.

TLA: Targeted locus amplification. A technique where regions of interest are amplified along with their ligated sequence determined via cross linking of physically proximal sequence.

WGS: Whole-genome sequencing. Short-read sequencing of an entire tumour or normal genome. Typically 30x coverage.

Figure Legends

Figure i - Power curves for detecting a copy-number change of. The y-axis is tumour purity (the fraction of tumour cells in the sample) and the x-axis is number of aligned reads. Any sample lying above the curve will have sufficient power to detect a copy-number change +1 with a bin size of 500kb. The percentage label next to each curve represents the percentage of tumour cells containing the copy-number change.

Figure ii - Chromosomal copy-number profile of a diffuse low grade glioma with a 1p/19q co-deletion generated by sWGS with DNA isolated from FFPE. Co-deletion of the entire 1p and 19q chromosomal arms are indicative of a watchful waiting treatment strategy. The y-axis represents normalized log2 sequence read counts per bin, and the x-axis represents 15 kb bins ordered by genomic position from chromosomes 1 to 22.

Figure 1 - A schematic of how different classes of structural variation can result in a potentially “actionable” genomic change. Alongside each is a summary of the sequencing and non-sequencing based tools that can be used for detection ordered from most comprehensive, to least. Where applicable, a prototypical example of a targetable mutation arising as a result of the class of structural variation is listed, along with its associated treatment. a) and c) are examples of balanced structural variants that do not result in any loss or gain of genetic material. In contrast, b) and d) are examples of unbalanced structural variants, also known as copy-number aberrations, that involve changes in the amount of genetic material in the nucleus. The technologies listed that interrogate all of these aberrations in a clinical setting involve a trade-off between cost, resolution, comprehensiveness, and applicability to formalin-fixed clinical samples. (Reviewed in [25,84,105].)