Liquid Biopsy Estimates of Tumour Mutational Burden to Predict Response to Immunotherapy Treatments for Cancer

Jacob Bradley[1] **and Nirmesh Patel PhD**[2]

[1]Student: Part III Systems Biology, University of Cambridge, UK
[2]Supervisor: Cambridge Cancer Genomics, Cambridge, UK

ABSTRACT

The field of immuno-oncology (IO) is exploding right now with most new oncology clinical trials focussed on IO and almost all pharmaceutical companies developing IO drug pipelines. The best predictor of treatment effectiveness in most cancers is the metric of tumour mutation burden (TMB) and micro-satellite instability (MSI) for colorectal cancers. Due to the signal to noise ratio problem when analysing ctDNA, this project aims to investigate the feasibility of calculating levels of TMB and MSI using our core technology pipelines, comparing this to TMB and MSI calculated using independent methods, and comparing solid with matched liquid biopsy samples. We would also look into the known IO resistance mechanisms so that we can ensure the best treatment monitoring possible. We expect is to use the results from this data to inform our decision on whether it is scientifically feasible to create a product to monitor treatment response in patients on IO drugs and to potentially create a IO treatment monitoring targeted panel.

Contents

1 List of Abbreviations **1**

2 Introduction **2**

3 Literature Review **2**

3.1 MSI . . . 2

3.2 TMB . . . 3

Relation to other biomarkers • Calculating TMB via whole exome sequencing • Efforts to calculate TMB via gene panels

3.3 Resistance Mechanisms to Immune Checkpoint Blockade . . . 4

Immunoregulatory Pathways • Metabolic Factors • Phagocytes • Microbiome • Hyperprogression • Unsustainable toxicity

4 Results **4**

5 Discussion **4**

6 Materials and Methods **4**

6.1 TMB Determination Via Whole Exome Data . . . 4

6.2 Candidate Loci Selection . . . 4

Optimising Sliding Window

References **5**

7 WORKING SECTION Paper Summaries **7**

Fabrizio et al. • Chen et al. • Sharma et al. • Chalmers et al. • Miller et al. • Vierira et al. • Koeppel et al. • Chang et al. • Stenzinger et al. • Dudley et al. • Lyu et al.

7.1 WORKING SECTION Potential Data Sources . . . 9

1 List of Abbreviations

- IO: Immuno-Oncology
- IT: Immunotherapy
- ICI: Immune Checkpoint Inhibitor

- TMB: Tumour Mutational Burden
- NAB: Neo-Antigen Burden
- MSI: Microsatellite Instability
- MSI-H: High Microsatellite Instability
- MSS: Microsatellite Stable
- MMR: Mismatch Repair
- dMMR: deficient Mismatch Repair
- ctDNA: Circulating Tumour DNA
- cfDNA: Cell-Free DNA
- tDNA: tumour DNA
- CTLA-4: Cytotoxic T-Lymphocyte Associated Protein 4
- B7: binds to CTLA-4 (also known as CD80)
- PD-1: Programmed Death Protein 1
- PD-L1: Programmed Death Ligand 1
- MHC: Major Histocompatibility Complex
- GEP: Gene Expression Profile
- DEG: Differentially Expressed Gene
- OS: Overall Survival
- CRC: Colorectal Cancer
- MM: Multiple Myeloma

2 Introduction

3 Literature Review

3.1 MSI

Microsatellites are regions of tandem repeats of short sequences distributed throughout the human genome (coding and non-coding regions). The sequences they are formed of tend to range in length between one and six bases, and tandem repeat sequences are typically 10-60 repeat units long. They are particularly prone to slippage during replication, resulting in new alleles differing in length (but not sequence) at a given microsatellite locus. This gives them a mutation rate several orders of magnitude higher than standard DNA, although this is highly non-uniform across chromosomes and species[1]. Their distinctive sequence pattern and high rate of mutation means they have been employed extensive as genetic markers in population studies to estimate genetic distance and linkage patterns[2] (particularly in plant studies). It has been possible to assess microsatellite status at a small number of loci without next generation sequencing technology via PCR amplification and capillary electrophoresis (measuring differences in the length of tandem repeat sequences).

Microsatellite instability is a term used to denote a high degree of variability (tandem repeat sequence length) in microsatellite regions within an individual- it is considered a marker of genomic instability. Microsatellite instability (MSI) was introduced in a clinical setting in the context of Lynch syndrome[3] (LS) and has been associated with deficiencies in mismatch repair machinery. Lynch syndrome is characterised by inherited germline mutations in genes coding for DNA mismatch repair machinery, and is associated with increased risk for a variety of cancers, but particularly colorectal cancer (CRC). CRC associated with LS has been associated with better prognosis but tentatively associated with poor response to chemotherapy[4] (although this remains controversial).

The improved prognosis of patients with MSI-High tumours is postulated to be associated with increased immune system response via increased presentation of neoantigen peptides which can be recognised by cytotoxic T-cells. Indeed, MSI-H tumours have been observed to have stronger presence of tumour infiltrating lymphocytes (TILs)[5]- although authors admit that this must happen in conjunction with increased cell death due to massive mutational burden compromising essential functions. Predictions about the relation between MSI and immune presence have led to sustained interest in microsatellite instability in the context of immunotherapy. Microsatellite instability has been shown in some cancers to be a strong indicator of response to immune checkpoint blockade[6]. To take an example, in colorectal cancer a progression free survival rate of 78% has been reported in association with MSI-H status, as opposed to 11% for microsatellite stable tumours[7]. The two immune checkpoint mechanisms currently targeted by checkpoint inhibition therapy (CTLA-4 and PD-1, targeted with ipilimumab and nivolumab respectively) have both had blockade antibodies approved by the USA's Food and Drug Administration in tandem with MSI as a companion diagnostic. While in many cancers MSI-H status is fairly rare or not seen at all (e.g. acute myeloid leukemia and uveal melanoma), in some cancers it is estimated to be present in 15-30% of cases[8] (e.g. colorectal and endometrial cancers). The increased response to immunotherapy even when controlling for typically improved prognosis in MSI-H cases lends credence to the hypothesis that MSI-H tumours are more strongly immunogenic that MSS tumours.

3.2 TMB

The immunogenicity of MSI-H tumours is proposed to derive from their increased neo-antigen burden. In this scheme MSI-H status correlates well with neo-antigen burden because deficiency in mismatch repair processes is accompanied by genetic alterations throughout the genome (including coding regions), some of which will, post-translation, be broken down and presented on the cell surface and act as epitopes for cytotoxic T-Cells. It has therefore been postulated that the overall tumour mutational burden (TMB) of a tumour exome would be a better predictor of neo-antigen burden, and by extension a better biomarker for response to immunotherapy. It is important to note that TMB includes not just mutations found commonly in cancer, but mutations throughout the genome. Therefore, only with recent advances in next generation sequencing has the use of TMB as a practicable marker emerged.

3.2.1 Relation to other biomarkers

It seems reasonable to assume that cancers which are microsatellite unstable will have high mutational burden. Whilst in many cases this appears to be true[9], it has not been possible to completely dismiss the alternative. What has certainly been shown that the relation between microsatellite status and TMB is very much dependent upon cancer type- the rare cases of MSI-H/TMB-L combination are almost exclusively found in a small group of cancers[10]. There are also cancers for which TMB-H status is very common but MSI-H is exceedingly rare- notable examples include melanomas and lung cancers, for which it is interesting to note that a larger portion of their mutation burden typically arises via exogenous factors (e.g. tobacco smoke and UV), and smoking has been associated with higher tumour mutational burden in cases of NSCLC[11]. In conclusion, the approximate assumption that MSI-H cases are a subset of TMB-H cases is not entire, but nor is it unreasonable with caveats (in particular deviation between cancers).

Given that TMB's promise lies in its prediction of neoantigen burden, it seems important to further investigate the relationship between the two. Measuring neoantigen burden directly is limited by modern methods, and must be established via analysis of exome sequene and computational modelling to estimate the binding affinity of short chain peptide fragments with cells' major histocompatibility complexes[12]. Studies into this utilise the same genome sequencing data as has been used to calculate tumour mutational burden, so it is not unsurprising that there is correlation found between the two. That said, this is still typically the case[13].

Other biomarkers for response to immune checkpoint inhibition have attempted to gauge the inflammation of T-cells in the tumour microenvironment, in particular via analysis of transciptomic gene expression profiles (GEPs). It has been demonstrated that a GEP Score can give comparable specificity and sensitivity to TMB in predicting clinical outcomes. Further than this, GEP score and TMB score show very low (almost no) correlation, indicating that they are best employed in conjunction, giving remarkably good predictions of tumour response[14].

3.2.2 Calculating TMB via whole exome sequencing

Physical restrictions on WES calculations of TMB include cost, time, and small sample size (even in solid tumour samples this can be a limiting factor). Sequencing entire exomes at a reasonable depth is a major undertaking. In particular blood based WES has suffered from poor recapitulation of variants detected in paired solid tumour samples, whilst amongst good ctDNA samples ctDNA estimated TMB correlated very well with tDNA-est TMB[15].

Beyond physical restrictions, there is very little consistency despite multiple effort to standardise the bioinformatic processes by which TMB is calculated[16]. Points of debate for mutations to include in TMB calculation stratify a range of features, including confidence in somaticmutation presence/sequence depth, allelic frequency (which may have an entirely different profile in blood biopsy), heterozygous vs homozygous frequency, type of mutation and protein effect[15]. There is also debate about whether mutations in known cancer driver genes should be included, as they are likely to experience a selection bias for the samples sequenced. In practice alternate approaches have tended to be highly correlated[17], so that the issue has been one of calibration and thresholding: this is encouraging, but currently means critical TMB scores have had little mutual relevance between assays and bioinformatics pipelines.

3.2.3 Efforts to calculate TMB via gene panels

There has been extensive disagreement about the viability of gene panels for calculating TMB with sensitivity and specificity. Most work done has been exploratory and acknowledges the need for validation in larger clinical trials[18]. Panels/models have been proposed by Zhuang et al.[19], Cristescu et al.[14], Lyu et al.[20], Johnson et al.[21] and Nichol et al.[22]. Ideally all of these should be more thoroughly reviewed before/while constructing a targeted gene panel.

3.3 Resistance Mechanisms to Immune Checkpoint Blockade

Need to write summary for these things

3.3.1 Immunoregulatory Pathways
3.3.2 Metabolic Factors
3.3.3 Phagocytes
3.3.4 Microbiome
3.3.5 Hyperprogression
3.3.6 Unsustainable toxicity

4 Results

5 Discussion

6 Materials and Methods

6.1 TMB Determination Via Whole Exome Data

We calculate TMB using whole exome data. Across TCGA and other resources there are a variety of methods utilised which have different genome coverages. We are looking to get a better grip on these but for the moment quote standard exome size as 30MB.

We only consider as somatic mutations:

- Frameshift Deletions
- Frameshift Insertons
- In-Frame Deletions
- In-Frame Insertions
- Missense Mutations
- Nonsense Mutations
- Nonstop Mutations
- Splice Sites
- Translation Start Sites

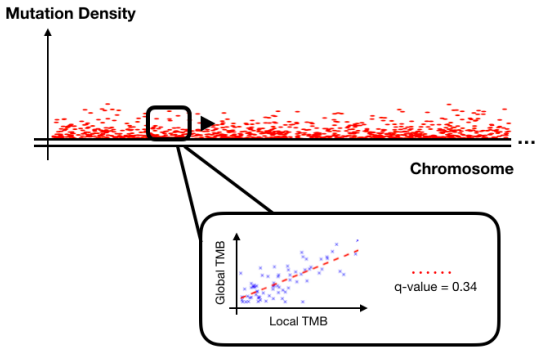
and we ignore the following 'silent' variants:

- 3'Flanks
- 3'UTRs
- 5'Flanks
- 5'UTRs
- IGRs
- Introns
- RNAs
- Splice Regions

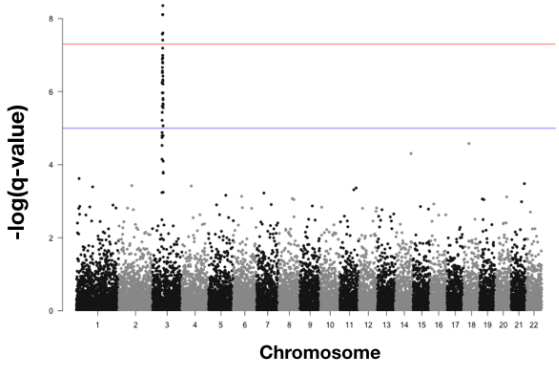
6.2 Candidate Loci Selection

In order to create an IO gene panel (or at least establish a pool of candidate genetic locations), we need to know which genome regions are most strongly associated with TMB. To determine regions whose mutation density is strongly indicative of overall burden, we want to use a sliding window algorithm (Figure 1a). This would involve:

1. For a codon, considering an interval of some width around it in the genome.
2. For this window, calculating the local mutation burden (for each patient separately).



(a) Sliding window correlates local mutation burden with global mutation burden.



(b) Manhattan plot showing regions of high association (example figure from R package qqman documentation[23]).

Figure 1. Genome-wide association with TMB using a sliding window algorithm.

3. Calculating a correlation metric across patients between the local burden at the interval around that codon, and the patient's whole exome TMB score. Converting this metric to a q-value.

4. Moving along one codon.

From this data we would be able to conduct genome-wide association analysis of regions significantly associated with TMB (Figure 1b). The baseline method for estimating TMB would be to take the overall mutation density of the genome regions selected, although in the long term better methods of calculation might emerge (e.g we might wish to weight the relative contributions to overall TMB as some gene regions may have a stronger sway, as has been attempted previous in canccer-specific gene panels[20]). Other factors to consider include whether synonymous mutations should be counted when attempting to estimate TMB. Synonymous mutations are typically excluded from global TMB calculations, but including them in a gene panel would give more data to use and therefore possibly lower variance.

We might not have time or resources, but I personally would be interested to see if there were non-coding regions significantly associated with TMB (e.g. 3'/5' UTRs).

6.2.1 Optimising Sliding Window

A problem we're currently considering is optimisation of the 'window size' metaparameter. Too small and the mutation rate would mean the window was unlikely to include enough mutated points to be viable, too large and granular information might be missed. Ideally we want to choose window size in an unbiased and optimal manner- we are looking at establishing a metric to gague the success of the sliding window algorithm depending on window size, that can then be used to tailor our approach.

Acknowledgements

References

1. Ellegren, H. Microsatellite mutations in the germline. *Trends Genet.* **16**, 551–558, DOI: 10.1093/mutage/ger028 (2000).
2. Vieira, M. L. C., Santini, L., Diniz, A. L. & Munhoz, C. d. F. Microsatellite markers: What they mean and why they are so useful. *Genet. Mol. Biol.* **39**, 312–328, DOI: 10.1590/1678-4685-GMB-2016-0027 (2016).
3. Hampel, H., Frankel, W. L., Martin, E. & Arnold, M. New England Screening for the Lynch Syndrome (Hereditary Nonpolyposis Colorectal Cancer). 1749–1759 (2005).
4. Jo, W.-S. & Carethers, J. M. Chemotherapeutic implications in microsatellite unstable colorectal cancer. **143**, 951–959, DOI: 10.1017/S0950268814002131.Tuberculosis (2015). 15334406.
5. Dolcetti, R. *et al.* High prevalence of activated intraepithelial cytotoxic T lymphocytes and increased neoplastic cell apoptosis in colorectal carcinomas with microsatellite instability. *Am. J. Pathol.* **154**, 1805–1813, DOI: 10.1016/S0002-9440(10)65436-3 (1999).

6. Dudley, J. C., Lin, M. T., Le, D. T. & Eshleman, J. R. Microsatellite instability as a biomarker for PD-1 blockade. *Clin. Cancer Res.* **22**, 813–820, DOI: 10.1158/1078-0432.CCR-15-1678 (2016).

7. Le, D. T. *et al.* PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med* **372**, 2509–2520, DOI: 10.1056/NEJMoa1500596.PD-1 (2016).

8. Bonneville, R. *et al.* Landscape of Microsatellite Instability Across 39 Cancer Types. *JCO Precis. Oncol.* 1–15, DOI: 10.1200/PO.17.00073 (2017). 15334406.

9. Fabrizio, D. A. *et al.* Beyond microsatellite testing: Assessment of tumor mutational burden identifies subsets of colorectal cancer who may respond to immune checkpoint inhibition. *J. Gastrointest. Oncol.* **9**, 610–617, DOI: 10.21037/jgo.2018.05.06 (2018).

10. Chalmers, Z. R. *et al.* Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Medicine* **9**, 1–14, DOI: 10.1186/s13073-017-0424-2 (2017).

11. Davis, A. A. *et al.* Association of tumor mutational burden with smoking and mutation status in non-small cell lung cancer (NSCLC). *J. Clin. Oncol.* **35**, 24, DOI: 10.1200/JCO.2017.35.7_suppl.24 (2017).

12. Gubin, M. M., Artyomov, M. N., Mardis, E. R. & Schreiber, R. D. Tumor neoantigens: building a framework for personalized cancer immunotherapy. *J. Clin. Investig.* **125**, 3413–3421, DOI: 10.1172/JCI80008.Introduction (2015).

13. Miller, A. *et al.* Correlation Between Somatic Mutation Burden, Neoantigen Load and Progression Free Survival in Multiple Myeloma: Analysis of MMRF CoMMpass Study. *Blood* **128**, 193 (2016).

14. Cristescu, R. *et al.* Pan-tumor genomic biomarkers for PD-1 checkpoint blockade–based immunotherapy. *Science* **362**, eaar3593, DOI: 10.1126/science.aar3593 (2018).

15. Koeppel, F. *et al.* Whole exome sequencing for determination of tumor mutation load in liquid biopsy from advanced cancer patients. *PLoS ONE* **12**, 1–14, DOI: 10.1371/journal.pone.0188174 (2017).

16. Stenzinger, A. *et al.* Tumor Mutational Burden (TMB) Standardization Initiative: Establishing a Consistent Methodology for TMB Measurement in Clinical Samples. *Int. J. Cancer* 2018, DOI: 10.1002/ijc.31878 (2018).

17. Chang, H. *et al.* Toward the Standardization of Bioinformatics Methods for the Accurate Assessment of Tumor Mutational Burden. 2018, DOI: 10.1136/esmoopen-2016-000114.30. (2018).

18. Steuer, C. E. & Ramalingam, S. S. Tumor Mutation Burden: Leading Immunotherapy to the Era of Precision Medicine? *J. Clin. Oncol.* **36**, JCO.2017.76.877, DOI: 10.1200/JCO.2017.76.8770 (2018).

19. Zhuang, W. *et al.* The tumor mutational Burden of Chinese advanced cancer patients estimated by a 381-cancer-gene Panel. *J. Cancer* **9**, 2302–2307, DOI: 10.7150/jca.24932 (2018).

20. Lyu, G. Y., Yeh, Y. H., Yeh, Y. C. & Wang, Y. C. Mutation load estimation model as a predictor of the response to cancer immunotherapy. *npj Genomic Medicine* **3**, 1–9, DOI: 10.1038/s41525-018-0051-x (2018).

21. Johnson, D. B. *et al.* Targeted next generation sequencing identifies markers of response to PD-1 blockade. **4**, 959–967, DOI: 10.1158/2326-6066.CIR-16-0143.Targeted (2017).

22. Nichol, D. *et al.* Pan-Cancer assessment of tumor mutational burden using a comprehensive genomic profiling assay. *J. Clin. Oncol.* **36**, 157, DOI: 10.1200/JCO.2018.36.5_suppl.157 (2018).

23. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The Am. J. Hum. Genet.* **81**, 559–575, DOI: 10.1086/519795 (2007). arXiv:1011.1669v3.

24. Chen, P.-l. *et al.* Analysis of immune signatures in longitudinal tumor samples yields insight into biomarkers of response and mechanisms of resistance to immune checkpoint blockade. **6**, 827–837, DOI: 10.1158/2159-8290.CD-15-1545.Analysis (2017).

25. Sharma, P. *et al.* Primary, Adaptive and Acquired Resistance to Cancer Immunotherapy. **168**, 707–723, DOI: 10.1016/j.cell.2017.01.017.Primary (2018).

26. Miller, A. *et al.* High somatic mutation and neoantigen burden are correlated with decreased progression-free survival in multiple myeloma. *Blood cancer journal* **7**, e612, DOI: 10.1038/bcj.2017.94 (2017).

7 WORKING SECTION Paper Summaries

I'm going to summarise the papers I've encountered so far, their main points and any shortcomings.

7.0.1 Fabrizio et al.

[9] *Beyond microsatellite testing: Assessment of tumor mutational burden identifies subsets of colorectalcancer who may respond to immune checkpoint inhibition.*

This paper aimed to increase the pool of people who, via easily ascertained metrics, might be shown likely receptive to immonutherapy for their cancer (colorectal). It's already been acknowledged that if you have high MSI you are more likely to respond to immune checkpoint therapy (78% versus 11% for MSS). This is supposedly because MSI is representtive of increased neo-antigen burden, the real driver of immune response. However, there seems a larger group of people, those who are TMB-high, who might be receptive to treatment. TMB status is ascertained via genomic profiling (in this case of solid tumour tissue), and is shown early to be very good at capturing MSI-H samples: almost everything that is MSI-H is TMB-H. However, about a third of TMB-H samples are MSS, so if they can show that TMB-H samples are likely receptive to IT then they've done good. They don't provide lots of direct evidence that TMB-H status increases likelihood of response to IT, but they do show that TMB-H status is significantly related to a variety of suspected/demonstrated oncogenic gene mutations. The other point they try to make is that their method of estimating TMB is good enough at the values of TMB seen in clinical practice.

7.0.2 Chen et al.

[24] *Analysis of immune signatures in longitudinal tumor samples yields insight into biomarkers of response and mechanisms of resistance to immune checkpoint blockade*

In this paper they analyse the utility of a few biomarkers for immune response, in particular the stages of treatment at which the biomarkers will reliably indicate responsiveness. They use CD8 and CD4 which are standard markers of T-cell density, and PD-L1, the binding ligand for ICI using PD-1, to indicate immunomodulation. They make the point that often these biomarkers will be modestly associated with responsiveness in the pre-treatment stage, but that the correspondence becomes far more striking (and significant) in early on-treatment samples. They highlight the importance of the invasive tumour margin, something that might be important (and unavailable) when dealing with liquid biopsy. They then measured gene expression profiles for a custom panel of genes to do with cancer signalling and/or immune response. They found no significant DEGs through pre-treatment and treatment with CTLA-4, but on early treatment of PD-1 blockade therapy they identified 411 DEGs, most of which were up-regulated in responders. However, in a different model which incorporated time trend analysis (in the form of a linear mixed effects model) they found many more differentially expressed genes.

7.0.3 Sharma et al.

[25] *Primary, Adaptive and Acquired Resistance to Cancer Immunotherapy*

Paper summarising knowledge of resistance mechanisms. Primary vs acquired resistance refers to the time after treatment at which resistance is observed. Adaptive refers to the mechanisms by which tumours dodge ICB, and can manifest as primary or acquired resistance. They discuss the mechanism of checkpoint, e.g. the competitive inhibition of CD28-B7 T-cell activation by CTLA-4. CTLA-4 also inhibits IL-2, an interleukin cytokine that stimulates T cell proliferation. Systemic steroid therapy is used to manage immune-related toxicity. The IFNgamma pathway is introduced. They discuss the possibilities of adoptive cell transfer, where T cells are engineered that specificially recognise tumour antigens. Tumours often escape via downregulating their MHC surface expression (also true in ICB).

7.0.4 Chalmers et al.

[10] *Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden*

This paper discusses further techniques for simplified computation of TMB (still from solid biopsy), its correlative relationship with MSI and particular mutational patterns in one gene heavily associated with TMB. They indicate that TMB can be estimated with reasonable accuracy from assessment of a far smaller gene panel than full exome sequencing- although this is more unreliable in small TMB samples (less important; we are looking for TMB-High). TMB distribution is highly *cancer-type dependent*, as is the specific relationship with TMB and MSI (although there are very few cancers in which TMB-Low+MSI tumours are common). TMB is *widely ranged* in all cancers. Unsurprisingly, TMB is associated strongly with DNA polymerase and mismatch repair LoF mutations. They point out that important mutations can occur in non-coding regions, particular if an upstream promoter/repressor is affected.

7.0.5 Miller et al.

[26] *High somatic mutation and neoantigen burden are correlated with decreased progression-free survival in multiple myeloma*

The point of this paper is that TMB and associated NAB are correlated with poor disease survival in an example of a cancer that on average has low TMB (Multiple Myeloma). This feeds into a question I have in general about mutational burden: you'd imagine in nastier cancers burden would be higher, so is it natural that people would respond more to therapy? Single agent IT

has actually been poor for OS in MM.

7.0.6 Vierira et al.

[2]*Microsatellite markers: what they mean and why they are so useful*
A vaguely helpful compilation of definitions and methods of calculation for microsatellite status.

7.0.7 Koeppel et al.

[15]*Whole exome sequencing for determination of tumor mutation load in liquid biopsy from advanced cancer patients*
In this paper the authors attempt to estimate TMB from ctDNA. Notably they do this via WES rather than a targeted gene panel, meaning their read depth is relatively low compared to what could potentially be achieved. They report that in cfDNA-WES samples that recapitulate variants called in cfDNA-TGS, TMB estimates correlated strongly between tDNA and cfDNA sequences (is this surprising?). They also identify useful considerations when estimating TMB, such as confidence in somatic mutation presence, sequence depth, allelic frequency (they also highlighted the discrepancy between relevance of the term allelic frequency in solid tumour versus plasma biopsy), allele count, mutated allele frequency in constitutional tissue, heterozygous and homozygous frequency, consequence on protein (missense, nonsense, inframe, frameshift...). The big problem identified in this study is very low quantities of sample (10ng)- again, focussing on a select gene panel could make better use of this.

7.0.8 Chang et al.

[17]*Toward the Standardization of Bioinformatics Methods for the Accurate Assessment of Tumor Mutational Burden*
This poster sets out a proposed workflow for standard calculation of TMB. They assume use of solid biopsy samples so use blood biopsy as germline representation. The tumour and liquid WES data are compiled separately, then co-aligned to a human reference genome. Germline mutations are subtracted, then some variants are filtered (they use SnpEff, dbSNP, ExAC, COSMIC, 1000 genomes project). They then classify different variants and sum the total. The paper compares different methods for computing TMB on a cohort. Missense only vs all mutations had a very high correlation (but a 3-fold factor difference). They note that indel calling algorithms can have impact on TMB calculation. A good germline genome is also important. Overall they note that there are offsets between different TMB calculation methods, but almost all are extremely highly correlated. The issue therefore becomes one of calibration and thresholding.

7.0.9 Stenzinger et al.

[16]*Tumour Muutational Burden (TMB) Standardization Initiative: Establishing a Consistent Methodology for TMB Measurement in Clinical Samples*
This poster gives several recommendations for reliable assessment and standardisation of TMB. They range from pre-analytical to post-result procedures. In selecting gene panels they recommend panels that screen for actionable mutations (unsure what specifically they mean?). They recommend panels with coverage of 1Mb or over (unclear where they get this number from?). They advise when constructing a panel to align TMB values to a WES-derived reference (unsure how else you would do it).

7.0.10 Dudley et al.

[6]*Microsatellite Instability as a Biomarker for PD-1 Blockade*
This paper introduces MSI as a biomarker associated with improved response to ICB, and briefly discusses distributon of MSI across patients and cancers. Mentions lymphocyte infiltration as a previous biomarker of good prognosis and response to immunotherapy. It links all the good biomarkers identified to mismatch repair deficiency causing microsatellite instability (so is the principle thing that generates variant allelism in MSI mismatch repair slippage?). They link increased immune response with neoepitope formation. They note that tumours with MSI had significant upregulation of PD-1 and PD-L1 (unsure whether they mean upregulation caused by dMMR, or that with dMMR the only tumours that survive have somehow up-regulated PD-1/L1). They point out an interesting feature that often PD-L1 is actually expressed on other T helper cells or other cells in the microenvironment, rather than tumour cells. MSI tumours actually tend to arise in different locations (at least in colorectal cancer) and have different appearances under the microscope. MSI has been used as a prognostic tool in some cases, although it's not entirely clear (also seems to be evidence MSI tumours actually do worse against chemotherapy). The paper discusss different criteria used to define MSI (in particular Lynch's syndrome) in the past. Bethesda guidelines, amsterdam criteria, NCCN guidelines. In the past MSI has been judged at a small sample of loci by amplifying segments via PCR and measuring their size via capillary electrophoresis. It's also been measured via immunohistochemistry for MMR proteins. Current guidance is a fairly long and convoluted progression of the different tests.

7.0.11 Lyu et al.

[20]*Mutation load estimation model as a predictor of the response to cancer immunotherapy* These guys have a small and 'clever' model for lung adenocarcinoma TMB- only 25 genes! Drawback is they reckon they have evidence that their model isn't extensible to other cancer types (not even enough genes in the panel to do the job for melanoma or colorectal). They use a linear model and a bayes information criteriion method for model selection (read over your third year coursework). It actually performs pretty well and their classification of response vs non-response is comparable to previous WES studies. They demonstrate that the model they've selected is better than almost all random models of the same size (you'd hope!!). They have some plot types that might be useful when deciding how to present the ideal. They apply the model trained purely on lung adenocarcinoma, and the performance isn't nealy as good (though better than nothing). Interestingly, while the lung adenocarcinoma model wasn't nearly as good at estimating TMB in melanoma as the melanoma model, they performed similarly well when reporting patient outcomes. This is worth bearing in mind when considering verification via WES TMB vs Clinical Outcomes.
END OF WORKING SECTION

7.1 WORKING SECTION Potential Data Sources

- TCGA
- TARGET database
- CheckMate Clinical Trials
- Keynote Clinical Trials
- NIH Clinical Genomic Database
- Texas Cancer Research Biobank
- COSMIC database

END OF WORKING SECTION