Preliminary exam proposal

Evolutionary and clinical characterization of a recurrent truncating frameshift in the HLA-A gene in microsatellite unstable tumors

Specific aims

Immunotherapy has emerged as an effective treatment option for a wide variety of cancers. Unfortunately, despite its successes, immunotherapy responders are still in the minority, with only an estimated 12% of patients having a durable response in 2018. Research is ongoing into the mechanisms through which cancer evades the immune system following immunotherapy. One well known mechanism of immune escape is the loss of function of one or more of the HLA genes. The HLA plays a key role in T-cell mediated cytotoxicity, which is the primary immune pathway exploited by many immunotherapies. Following loss of HLA function, T-cells become unable to identify cancer cells, and T-cell immunity is abolished. The link between loss of HLA in cancer and poor prognosis has been known for decades, but research in this area has largely been restricted to immunohistochemical studies. Attempts to investigate the genetics underpinning the loss of HLA in cancer are hindered by the fact that the HLA genes are extremely polymorphic, with thousands of alleles present in the population. Classic methods for genetic analysis rely on alignment of sequencing reads to a standard reference genome, which fails for the HLA genes where the average person is highly diverged from the standard reference. To tackle this problem, we have created Polytect, a tool that allows us to investigate somatic mutations in the HLA genes by dynamically constructing a personalized HLA reference which can then be used for alignment and variant calling.

Our long-term goal is to use Polytect to provide a better understanding of the specific genetic mechanisms responsible for the loss or alteration of HLA function in cancer. In this proposal, I will focus on the loss of HLA in cancers with microsatellite instability (MSI) where we have observed a frameshift mutation hotspot at a characteristic cytosine homopolymer at the beginning of exon 4 of the HLA-A gene (p.186fs). Our central hypothesis is that there is positive selection for these frameshift mutations in immune challenged tumors, and that the likelihood of this mutation can be predicted by an individual's HLA genotype.

Aim 1: Determine if there is positive selection for p.186fs mutations in immune challenged tumors

Tumors with MSI are generally considered to have a favorable prognosis due to their high mutational burden, and therefore high neoantigen load, which facilitates a positive immune response. This leads us to hypothesize that there should be a selective advantage for MSI tumors that lose HLA function due to a p.186fs mutation.

Aim 1a. Leveraging existing paired tumor/normal DNA sequencing data from sources such as TCGA, we will use a statistical model comparing the frequency of p.186fs mutations to those of neutral microsatellites to determine if there is positive selection for mutations at this hotspot in MSI tumors.

Aim 1b. Using RNA sequencing data from the same patients, we will compute immunophenoscores for each individual and determine if the strength of this selection is associated with the level of immune activity within the tumor.

Aim 2: Determine if patient HLA-A genotype predicts response to immunotherapy

MSI is known to lead to indels in homopolymers with a frequency that is exponentially increasing with the length of the homopolymer. We have observed that HLA-A alleles in the general population are split relatively evenly into two groups based on whether their cytosine homopolymer at the p.186fs hotspot is of length 5 (HLA-A-5C) or length 7 (HLA-A-7C). We therefore hypothesize that patients with an HLA-A-7C allele are at an increased risk of developing a p.186fs mutation, and that this will negatively affect survival following immunotherapy.

Aim 2a. We will use longitudinal sequencing data from patients with MSI tumors who have undergone immunotherapy to determine if HLA-A-7C alleles are at greater risk of developing a p.186fs mutation following treatment.

Aim 2b. Using survival data from the same cohorts, we will calculate hazard ratios to determine if p.186fs significantly affects survival and if HLA-A-7C is predictive of poor survival following immunotherapy.

Background and preliminary data

T-cell mediated cytotoxicity is one of the major pathways through which the immune system is able to eliminate malignant cells[1]. T-cells identify cells for elimination based on the presence of small peptides that are not found in the germline proteome and are therefore interpreted as foreign and harmful. Since most proteins are found on the inside of cells, where T-cells do not have access, they rely on the human leukocyte antigen (HLA) proteins to present intracellular proteins on the cell surface. The HLA proteins are constantly being loaded with peptide fragments produced from all proteins inside the cell, which are then transported to the cell surface where they can be detected by T-cells. In a healthy cell, the cell surface will be covered in HLA proteins presenting “self” peptides that do not provoke a T-cell response. However, whenever a cancer cell develops a coding region mutation that produces a novel protein sequence, it has been shown that these novel sequences can be presented by the HLA proteins and targeted for elimination by T-cells[2]. These novel sequences are termed cancer antigens or neoantigens, and the interaction between these antigens and the HLA form the foundation of the immune response against tumors. When the system is working, neoantigens are presented by the HLA proteins and tumor cells are eliminated. However, the fact that cancer exists is proof that this process does not always work, and as a result there is significant research looking into how cancer cells can evade T-cell mediated cytotoxicity.

There are many ways that cancer can evolve to evade the immune system. One of the early mechanisms identified was the expression of programmed death-ligand 1 (PD-L1), a ligand for the T-cell inhibitory receptor PD1[3]. In PD-L1 expressing tumors, T-cells will successfully identify tumor cells as malignant but will not actually kill the target cell due to PD1 mediated inhibition. As a result, anti-PD1 and anti-PD-L1 therapies were developed that block this inhibition and allow T-cells to kill their targets. In general, this is how immunotherapies work - they identify an evasive mechanism and disable it. However, immunotherapies rely on the assumption that neoantigens are being presented by the HLA proteins. If HLA presentation is eliminated, then no amount of T-cell focused immunotherapy will help since T-cells will never be able to find their targets.

Unsurprisingly, loss of HLA expression itself is an attractive immune escape mechanism for tumors. Immunohistochemical studies have shown that reduced HLA expression is common across cancers of different tissues [4], and this loss of expression has been linked to reduced patient survival [5]. We know that reduced HLA expression can be caused by reversible mechanisms such as transcriptional downregulation, which can be treated with drugs like Interferon-γ that increase HLA transcription, or by irreversible mechanisms such as truncating frameshifts, which are essentially untreatable short of gene therapy. Being able to identify irreversible genetic loss of HLA expression is valuable so that we can identify patients that are likely immunotherapy non-responders and avoid putting them through treatment regimens that are unlikely to work.

Identification of the specific genetic mechanisms underpinning loss of HLA expression requires a move from immunohistochemical studies to DNA sequencing of tumor cells. This is difficult for the HLA genes due to their unusually high polymorphism within the human population. There are three HLA class 1 genes involved in the presentation of peptides (HLA-A, HLA-B, and HLA-C) and each is currently known to have over 5000 alleles coding for over 3000 different proteins[6], with hundreds of new alleles still being discovered every year. In traditional genetic studies, DNA is sequenced from paired tumor and normal samples, reads are aligned to the human reference genome, and differences between the tumor and normal are called as somatic mutations. In the case of the HLA genes, due to high polymorphism the average person is highly diverged from the human reference genome with an average of 40 single nucleotide variants within the coding regions of each allele (Figure 1). Since the average person is also heterozygous with two divergent alleles at each locus this is an

average of 80 coding variants when aligning to the haploid standard reference, or a variant every ~15 bases. To complicate matters further, there are 3 non-classical HLA genes and 12 pseudogenes that all have high homology to the classical HLA class 1 genes[6]. This genetic complexity at the HLA locus causes poor alignments following DNA sequencing, which prevents high quality mutation calls. Some studies have identified irreversible genetic loss of HLA expression due to large structural variants such as loss of heterozygosity at the HLA locus[7], or mutations in the non-polymorphic Beta-2-M protein which is required for HLA complex formation[8]. However, small variants in the HLA coding regions have been understudied.

To address this problem, we created Polytect, a genomics tool that can dynamically construct a set of personalized HLA reference sequences to use for alignment and mutation calling. The key idea behind Polytect is that given a specific germline allele, if you were to create a set of reads from that allele and align them to another allele you would get a specific pattern. For example, if you were to perform 10 sequencing runs generating reads from allele HLA-A*01:01:01:01, these reads would align to allele HLA-A*02:01:01:01 in a similar pattern every time. We take advantage of this by taking a database of known alleles for each HLA gene, simulating reads from each allele, and aligning them to each other allele in the database. We then model this alignment as a linear system that lets us solve for the underlying diploid haplotype. Our solution gives us the most likely germline sequences present within our database, but it is possible to have an individual with private variants, or a slightly incorrect call by the algorithm. To improve our references, we then perform the additional step of alignment of normal (non-tumor) reads to our inferred references followed by imputation of any apparent germline mutations into the reference to create a truly personalized reference. Initial benchmarks using a gold standard truth set show that Polytect creates reference sequences with a median Levenshtein distance of 0 from the true underlying HLA sequences for all 3 HLA class 1 genes (Figure 1). With Polytect's reference sequences, we can now align our HLA reads and call phased somatic mutations using traditional methods.

Our initial goal with Polytect was to report the landscape of somatic HLA mutations in metastatic cancer, following up on an earlier study that looked at somatic mutations across primary cancers using data from the Cancer Genome Atlas (TCGA)[9]. We began by applying Polytect to TCGA samples to first see if we could reproduce their results. In the colon adenocarcinoma (COAD) cohort we observed a truncating frameshift hotspot at the beginning of exon 4 of HLA-A (p.186fs), an observation also reported by Shukla et al. However, while they noted the apparent recurrent mutation, they did not investigate further. When looking at the recurrence ourselves we noticed that it occurs mainly in colon adenocarcinomas, stomach adenocarcinomas, and endometrial cancers, all of which are known to have significant numbers of cases with the microsatellite instability (MSI) phenotype[10].

MSI is a hypermutated phenotype caused by impaired DNA mismatch repair[11]. During normal operation DNA polymerases can slip when moving along repetitive DNA sequences (microsatellites) which can introduce insertions or deletions into the DNA[12]. In the absence of DNA mismatch repair, these indels are retained leading to an accumulation of indels within microsatellites in MSI tumors. When looking at the genomic context of p.186fs, we noticed that it occurred at a cytosine homopolymer microsatellite, supporting the idea that this recurrence was driven by MSI. Further, since the COAD cohort is phenotyped for MSI, we noted that 100% of individuals (11 of 11) in this cohort harboring p.186fs were annotated as MSI high (MSI-H), despite the fact that the MSI-H phenotype is only present in 16.7% of the cohort (Table 1). Upon further investigation, we found that this microsatellite is polymorphic within the human population. All HLA-A alleles are split between whether their cytosine homopolymer at this location is of length 5 (HLA-A-5C) or of length 7 (HLA-A-7C), with each variant present at an allele frequency of 0.5 within the cohort. MSI is known to lead to indels in homopolymers at a rate exponentially increasing with the length of the homopolymer[13] (Figure 2), so we expected HLA-A-7C alleles to account for the majority of p.186fs mutations. Within the COAD cohort, we observed 83% (10 of 12) of p.186fs mutations occurring within HLA-A-7C alleles (Table 2). Strikingly, this is an allele specific recurrence of 22.2% within MSI-H tumors.

MSI-H tumors normally have a favorable prognosis due to the fact that high mutation burden leads to high neoantigen load, which provides the immune system with more possible targets for T-cell mediated

cytotoxicity[14]. This led to the FDA granting accelerated approval of the anti-PD1 immunotherapy pembrolizumab for treatment of MSI-H tumors, the first case of FDA accelerated approval for immunotherapy based on a biomarker. However, based on this preliminary data, we hypothesized that a subset of MSI-H patients with HLA-A-7C alleles may experience decreased benefit from immunotherapy due to existing or increased risk of loss of HLA-A expression via p.186fs mutations. This proposal lays out a plan to investigate whether loss of HLA-A via p.186fs reduces patient survival, and if individuals with HLA-A-7C alleles are at increased risk of HLA-A loss following immunotherapy.

Significance: Irreversible loss of HLA expression prevents effective targeting of tumor cells by the immune system and invalidates many immunotherapies. Identification of a common recurrent mutation leading to irreversible loss of HLA-A in patients with MSI tumors will allow for more informed decisions when selecting treatment options for patients who already have the identified loss of HLA-A, or are at risk of developing it due to their HLA-A genotype.

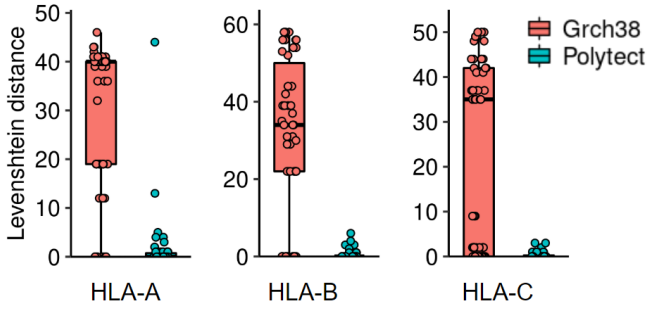

Figure 1. Levenshtein distance between the HLA-class 1 exons of Grch38 or Polytect inferred reference sequences and true underlying germline sequences as measured from a gold standard dataset of 29 individuals with known HLA haplotypes.

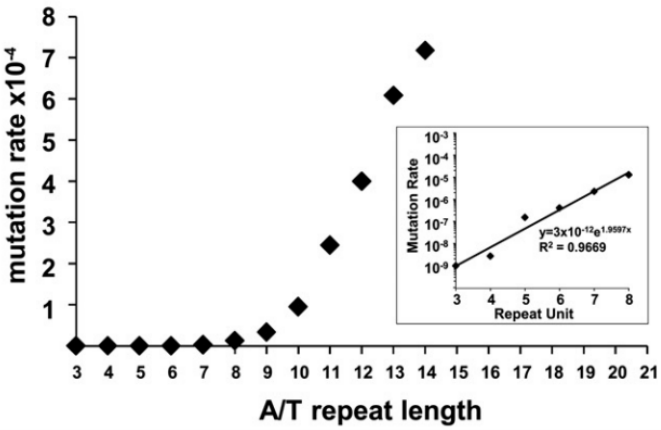

Figure 2. Reproduced from Lang et.al 2013. Mutation rate (mutations per repeat per generation) for homopolymer microsatellites plotted according to repeat length. The exponential increase for repeats of length 3 to 8 are plotted in log scale in the insert.

MSI status	COAD patients with phenotype	COAD patients with p.186fs	% patients with p.186fs

MS-stable	172	0	0.0%
MS-low	52	0	0.0%
MS-high	45	11	24.4%

Table 1. Presence of p.186fs across the TCGA colon adenocarcinoma (COAD) cohort by MSI status.

HLA-A allele	Alleles within MSI-H samples	p.186fs mutations	Recurrence
HLA-A-5C	45	2	4.4%
HLA-A-7C	45	10	22.2%

Table 2. Presence of p.186fs by homopolymer length.

Research strategy

Aim 1: Determine if there is positive selection for p.186fs in immune challenged tumors

Datasets: The Cancer Genome Atlas is a repository containing both sequencing and phenotype information for thousands of cancers across multiple cancer types. TCGA contains 83 colon adenocarcinoma (COAD), 85 stomach adenocarcinoma (STAD), and 170 uterine corpus endometrial carcinomas (UCEC) that have been phenotyped as MSI-H. This is a total of 338 MSI-H primary tumors.

Aim 1a. Comparing the frequency of p.186fs mutations to those of neutral microsatellites to determine if there is positive selection in MSI-H tumors.

Rationale: Tumors with MSI are generally considered to have a favorable prognosis due to their high mutational burden, and therefore high neoantigen load, which facilitates a positive immune response[15]. This leads us to hypothesize that there should be a selective advantage for MSI-H tumors that lose HLA-A function due to p.186fs resulting in positive selection of the mutation.

Research strategy: Positive selection of indels at a particular microsatellite can be detected as an enrichment of indels at the loci above what would be expected if that loci was a neutral microsatellite. A model that attempts to determine the probability of observing indels at a specific microsatellite needs to take into account both inter-tumor mutation rates which can vary by up to 3 orders of magnitude within MSI-H cancers, and intra-tumor mutation rates which can vary based on the length and composition of a given microsatellite.

We will perform simulations from a null model that assumes microsatellites are not under selection, followed by an analysis to see if the observed rate of indels at the location of p.186fs is higher than expected based on the simulations. Repeat composition and repeat length affect the rate of indels in microsatellites, so our null model will only consider indels in cytosine/guanine homopolymers of length 5 or 7. The number of indels per sample is heterogeneous between tumors and will be modeled as a Beta-binomial random variable. The parameters for the Beta-binomial random variable will be determined by fitting the distribution to the number of observed indels within known homocytosine/homoguanine homopolymers in our cohort. We will run 10,000 simulations, each with a number of samples equal to our experimental cohort. For each sample, the number of indels occurring will be drawn from the Beta-binomial distribution, and each indel will be assigned to

a microsatellite location uniformly randomly. Within each simulation we will count the number of p.186fs mutations. Significant enrichment will be determined by the presence of our observed counts being in the 95th percentile based on the simulations.

Pitfalls and alternative strategies: Our model assumes that the majority of microsatellites are neutral, but when using whole exome sequencing this is unlikely to be true. Previous studies have shown that in MSI tumors indels in coding region microsatellites appear to be under negative selection[16], suggesting that the null model's assumption of neutral selection for p.186fs could be too conservative. As an alternative approach, we can look at indels within entire genes rather than within particular microsatellites. Enrichment of truncating frameshifts within the HLA genes relative to other genes could also signal positive selection for loss of function mutations.

Aim 1b. Compute immunophenoscores for each individual and determine if the degree of positive selection is associated with the level of immune activity within a tumor.

Rationale: We hypothesize that p.186fs provides a selective advantage to tumors by allowing them to evade the immune system. We therefore expect that p.186fs should be selected for more often in tumors where there is high immune activity and higher selective pressure.

Research strategy: Immune activity in tissues can be quantified using immunophenoscores, a technique that combines expression levels of 162 immune system genes into a single score[17]. The immunophenoscore model accounts for genes related to infiltration of T cells into the tumor, infiltration of immunosuppressive cells into the tumor, expression of MHC molecules, and expression of co-inhibitory or co-stimulatory molecules. Immunophenoscores range from 0 to 10, with 0 representing the lowest amount of immune activity within the tumor. We will calculate immunophenoscores for every sample in our cohort and use logistic regression to determine if p.186fs occurs more often in high scoring tumors.

Aim 2: Determine if patient HLA-A genotype predicts response to immunotherapy

Datasets: Sufficient longitudinal sequencing data in MSI-H patients does not exist at this time. To gather this data, patients with MSI-H tumors will be followed longitudinally, with whole exome sequencing performed on normal tissue at the beginning of the study, and in tumors before and at 1, 3, 6, 9, and 12 months after the initiation of immunotherapy. Based on previous reports showing a 10-fold increase in rate of indels in 7-mer homopolymers vs 5-mer homopolymers (Figure 3, insert), power analysis for a Cox proportional hazards model shows that we would need 178 alleles (89 patients) to detect a significant increase in indel risk in HLA-A-7C alleles vs HLA-A-5C alleles with a 5% Type 1 error rate and power of 0.8.

Aim 2a: Use longitudinal sequencing data to determine if HLA-A-7C alleles are at greater risk of developing a p.186fs mutation following treatment.

Rationale: MSI is known to lead to indels in homopolymers with a frequency that is exponentially increasing with the length of the homopolymer. We have observed that HLA-A alleles in the general population are split relatively evenly into two groups based on whether their cytosine homopolymer at the p.186fs hotspot is of length 5 (HLA-A-5C) or length 7 (HLA-A-7C). We therefore hypothesize that HLA-A-7C alleles are at an increased risk of developing a p.186fs mutation compared to HLA-A-5C.

Research strategy: For this analysis, we will select only individuals from our cohort that are p.186fs negative before the initiation of immunotherapy, followed by survival analysis on the time to development of p.186fs. To determine the allele-specific risk of mutation, we will follow specific alleles through all time points, with each

patient contributing two alleles to the analysis. The Cox proportional hazards model will be used to determine if there is a significant difference in the risk of developing p.186fs in HLA-A-7C vs HLA-A-5C alleles.

Pitfalls and alternative strategies: The strength of the immune response following immunotherapy may be highly variable between subjects. This could confound the analysis if it results in varying levels of selective pressure for p.186fs. As an alternative approach, we could also perform RNA sequencing on subjects which would allow us to compute immunophenoscores to include as a covariate in the model.

Aim 2b: Calculate hazard ratios from survival data to determine if p.186fs mutations significantly affect survival, and if HLA-A-7C is predictive of poor survival following immunotherapy.

Rationale: Loss of HLA-A activity due to p.186fs will result in a reduction of the peptide repertoire presented to T-cells, which should reduce the efficacy of T-cell mediated immunotherapy. We therefore hypothesize that patients with p.186fs mutations will have reduced survival, and that HLA-A-7C haplotypes will be predictive of poor survival following immunotherapy due to increased risk of developing a p.186fs mutation.

Research strategy: Our cohort will be separated into two groups, those with and those without p.186fs mutations. Cox proportional hazard ratios will be used to determine whether survival time is significantly reduced in individuals with p.186fs mutations following immunotherapy. To account for individuals that begin the study without p.186fs, but then develop it at some point during the study, we will use the Branson and Whitehead parametric estimation procedure to account for patient group switching. Using the same statistical model, we will also group the patients by haplotype to determine if HLA-A-7C dosage is predictive of poor survival following immunotherapy.

Pitfalls and alternative strategies: For aim 2b, power analysis shows that given a treatment group of 89 patients and an expected 1 year death rate of 30%, we would only be able to detect differences in survival with 80% power and 5% Type 1 error rates with hazard ratios of 5 or higher, which would be a fairly large difference. However, the germline HLA-A-7C dosage analysis only requires a single germline sequencing event with longitudinal survival follow-up, and not sequencing at every time point. This data could be supplemented with data from other clinical studies that have this data as part of a meta-analysis to increase sample size and power. A meta-analysis with follow-up of 500 patients would be able to detect differences between groups with a hazard ratio as low as 2.7.

Expected outcomes and alternative explanations

In aim 1a, we expect to see that the recurrent p.186fs mutation appears at a higher frequency than would be expected from chance alone following the accumulation of random indels. In aim 1b, we expect to provide evidence for an explanation for this phenomenon - that immune pressure leads to positive selection of p.186fs, which ostensibly provides a mechanism of immune escape. If we do not find the frequency of p.186fs to be significantly elevated, or we do not find the presence of p.186fs to be associated with high immune cell activity, it is possible that the mutation itself does not provide enough of an advantage to be selected for. One explanation for this could be that there are 6 total HLA class 1 alleles (2 each for HLA-A, HLA-B, and HLA-C) and losing only 1/6th of HLA presentation capacity is not enough for immune escape.

In aim 2a, we expect to see that HLA-A-7C alleles are at ~10-fold higher risk of developing p.186fs than HLA-A-5C alleles simply due to length of the homopolymer repeat. If this hypothesis does not bear out, it would suggest that microsatellite instability is not the main driver of mutations at this locus, and perhaps there is another mechanism of action responsible for development of p.186fs in these tumors. Mutational signature analysis would be an appropriate follow-up to attempt to determine what other mutational process could be causing the apparent recurrence.

In aim 2b, we expect to see that p.186fs reduces survival in patients, and that germline HLA-A-7C dosage predicts survival following immunotherapy. If aim 1b and aim 2a show negative results, it is likely that aim 2b will also return negative results for similar reasons. However, if aim 1b and aim 2a show positive results, but aim 2b shows no differences, one explanation would be that immunotherapy is powerful enough to overcome any p.186fs-mediated immune escape such that those with and those without the mutation show equal improvement. Follow-up studies looking at recurrent loss of HLA-B and HLA-C alongside HLA-A would be necessary to see if more complete loss of HLA function would have an impact on survival.

References

1. Chen, D. S. & Mellman, I. Oncology meets immunology: the cancer-immunity cycle. *Immunity* **39**, 1–10 (2013).
2. Boon, T., Cerottini, J. C., Van den Eynde, B., van der Bruggen, P. & Van Pel, A. Tumor antigens recognized by T lymphocytes. *Annu. Rev. Immunol.* **12**, 337–365 (1994).
3. Latchman, Y. *et al.* PD-L2 is a second ligand for PD-1 and inhibits T cell activation. *Nat. Immunol.* **2**, 261–268 (2001).
4. Garrido, F. *et al.* Natural history of HLA expression during tumour development. *Immunol. Today* **14**, 491–499 (1993).
5. Speetjens, F. M. *et al.* Clinical impact of HLA class I expression in rectal cancer. *Cancer Immunol. Immunother.* **57**, 601–609 (2008).
6. Robinson, J. *et al.* The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Res.* **43**, D423–31 (2015).
7. McGranahan, N. *et al.* Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. *Cell* **171**, 1259–1271.e11 (2017).
8. Del Campo, A. B. *et al.* Immune escape of cancer cells with beta2-microglobulin loss over the course of metastatic melanoma. *International journal of cancer* **134**, 102–113 (2014).
9. Shukla, S. A. *et al.* Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat. Biotechnol.* **33**, 1152–1158 (2015).
10. Cortes-Ciriano, I., Lee, S., Park, W.-Y., Kim, T.-M. & Park, P. J. A molecular portrait of microsatellite instability across multiple cancers. *Nat. Commun.* **8**, 15180 (2017).
11. Jiricny, J. The multifaceted mismatch-repair system. *Nat. Rev. Mol. Cell Biol.* **7**, 335–346 (2006).
12. Wierdl, M., Dominska, M. & Petes, T. D. Microsatellite instability in yeast: dependence on the length of the microsatellite. *Genetics* **146**, 769–779 (1997).
13. Lang, G. I., Parsons, L. & Gammie, A. E. Mutation rates, spectra, and genome-wide distribution of spontaneous mutations in mismatch repair deficient yeast. *G3* **3**, 1453–1465 (2013).
14. Chang, L., Chang, M., Chang, H. M. & Chang, F. Microsatellite Instability: A Predictive Biomarker for Cancer Immunotherapy. *Appl. Immunohistochem. Mol. Morphol.* **26**, e15–e21 (2018).
15. Mlecnik, B. *et al.* Integrative Analyses of Colorectal Cancer Show Immunoscore Is a Stronger Predictor of Patient Survival Than Microsatellite Instability. *Immunity* **44**, 698–711 (2016).
16. Kim, T.-M., Laird, P. W. & Park, P. J. The landscape of microsatellite instability in colorectal and endometrial cancer genomes. *Cell* **155**, 858–868 (2013).
17. Charoentong, P. *et al.* Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell Rep.* **18**, 248–262 (2017).