Next-generation sequencing of human tumor xenografts is significantly improved by prior depletion of mouse cells

Permanent Abstract Number 1455

David Agorku[1], Stefan Tomiuk[1], Kerstin Klingner[2], Stefan Wild[1], Silvia Rüberg[1], Lisa Zatrieb[1], Andreas Bosio[1], Julia Schueler[2], and Olaf Hardt[1]
[1] Miltenyi Biotec GmbH, Bergisch Gladbach, Germany | [2] Oncotest GmbH, Freiburg, Germany

Introduction

Human tumor xenografts represent the gold standard method for many research areas, including drug discovery, cancer stem cell biology, and metastasis prediction. When compared to *in vitro* cell culture models, human tumor xenografts show a higher validity for most assays[1].
During the growth phase *in vivo*, xenografted tissue is vascularized and infiltrated by cells of murine origin. The level of infiltration is highly dependent on multiple factors like tumor subtype, growth rate, and region of transplantation. However, even when these factors are kept constant, the amount and composition of infiltrating mouse cells is highly variable. Due to this, molecular downstream analyses such as microarray-based expression profiling are biased by cross-hybridization of mouse-derived molecules to human probes. In addition, a reduction of sensitivity caused by measuring undesired mouse signals during next-generation sequencing analysis can be expected.
To overcome these limitations, we developed a fast and easy method allowing for the effective depletion of all cells of mouse origin by using automated tissue dissociation and magnetic cell sorting (MACS® Technology). We performed whole exome sequencing (WES) of bulk human tumor xenografts from lung, bladder, and kidney cancer, and compared the results to samples depleted of mouse cells.

Results

1 Reliable and fast depletion of mouse cells

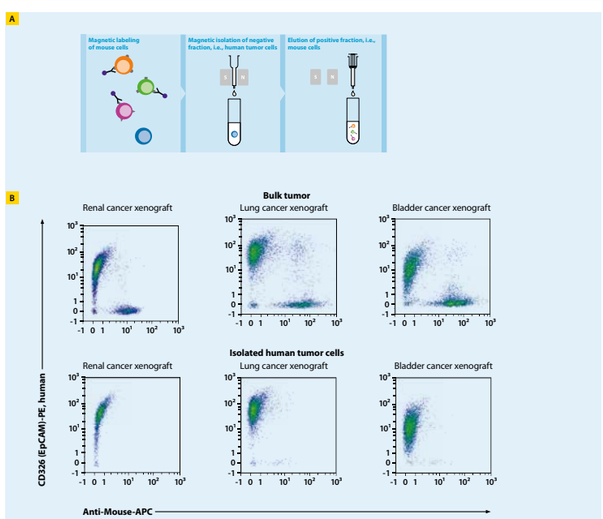

Figure 1

We have determined an antibody combination recognizing all mouse cells from all tissues. Conjugates of these antibodies with superparamagnetic nanoparticles (MicroBeads) were used to develop an optimized protocol for the depletion of mouse cells from human tumor xenografts by magnetic separation (fig. 1A).

It was possible to eliminate >99% of the contaminating mouse cells in less than 20 min, regardless of the tumor type. Cell fractions were labeled with the pan-mouse antibody cocktail and an antibody against human CD326 (EpCAM) (fig. 1B).

2 Whole Exome Sequencing of tumor samples prior to and after mouse cell depletion (MCD)

A

Xenograft	Description	Cluster density (K/mm²)	Estimated yield (MB)
Kidney	Bulk tumor	777	2841.8
Kidney	Isolated human tumor cells	1179	4195.8
Lung	Bulk tumor	793	2904.4
Lung	Isolated human tumor cells	1020	3678
Bladder	Bulk tumor	843	3063.6
Bladder	Isolated human tumor cells	1075	3859.2

B

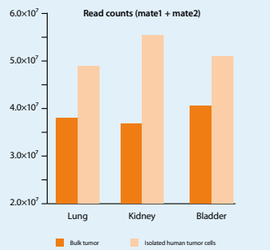

Figure 2

To assess the impact of MCD on the quality of next-generation sequencing data, we conducted WES on three different xenograft models derived from human kidney, lung, and bladder cancer.
As the capture oligonucleotides used for targeted enrichment of protein-coding sequences were designed based on the human genome, an initial pre-enrichment of DNA fragments of human origin from the mixture of mouse and human cells was expected. In order to assess the number of capture oligonucleotides that might cross-hybridize with mouse genomic DNA, we conducted BLAST searches of each single Nextera® probe against mouse genome and used the resulting alignment parameters to determine possible cross-hybridization. Depending on the selection thresholds (alignment length, no. of mismatches, no. of gaps), we predicted a cross-reactivity of 5–10% of capture probes with mouse transcripts (data not shown).
DNA from bulk tumor or isolated human tumor cells was used to produce exome-captured sequencing libraries applying the Nextera Rapid Capture Exome Kit (Illumina®). For sequencing on the MiSeq® instrument (Illumina) the MiSeq Reagent Kit v3 (150 cycles, Illumina) was utilized to generate 75-bp paired-end reads. A significant increase ($p < 0.05$) in cluster density (fig. 2A) as well as an average increase in read counts of 33% (fig. 2B) was observed for the samples depleted of mouse cells, indicating improved sample quality. Correspondingly, we observed a strong reduction of debris and dead cells upon MCD by flow cytometry analysis.

3 MCD strongly reduces number of erroneously mapped mouse reads

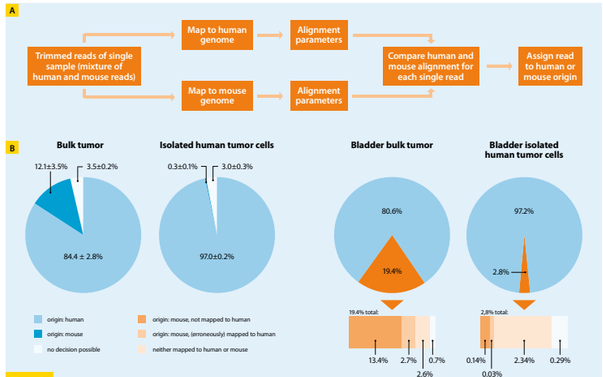

Figure 3

After adapter clipping (trimmomatic v0.32[2]), we mapped the reads of all samples against human and mouse genomes (bwa v0.7.12[3]) and determined their putative origin based on the respective alignment parameters (LINUX shell, command-line Perl) (fig. 3A). An average of 12% of reads derived from bulk tumor samples was attributed to mouse cells. This amount could be reduced to 0.3% by prior depletion of mouse cells (fig. 3B). As on average 15% of the mouse-derived reads mapped erroneously to the human genome (1.9% of total reads) in the bulk tumor samples, a strong positive influence of mouse cell depletion (0.04% of total reads erroneously mapped to human genome) on downstream analyses can be expected. Figure 3C exemplifies the detailed read assignment for bulk tumor and isolated human tumor cells derived from the bladder cancer xenograft.

4 MCD strongly reduces the number of falsely predicted SNPs

Figure 4

In order to determine the impact of mouse reads erroneously mapped to the human genome in the bulk tumor samples, we determined the number of predicted SNPs for the xenograft samples prior to and after MCD. As no healthy tissue was available for comparison, an SNP was defined as a difference between the sequenced sample and the reference genome (hg19). After removal of duplicate reads by MarkDuplicates (Picard Tools v.1.119[4]), SNP and INDEL calling was conducted using VarScan v2.3.7[5] and was restricted to the regions targeted by the Nextera Rapid Capture Exome Kit as provided on the Illumina homepage. 63±10% of all SNPs predicted for the bulk tumor samples were no longer detected after mouse cell depletion, 18±1% were specific for the isolated human tumor cells (fig. 4A). While the former were mainly caused by erroneously mapped mouse reads, the latter seemed to be detected due to higher read counts and accordingly higher coverage within the isolated human tumor cell samples. This effect was also visible for predicted INDELs (fig. 4B).

5 MCD improves the prediction of high-impact SNPs

Figure 5A exemplifies the impact of MCD on the prediction of a protein-coding exon of the POLA1 gene (generated by using the Integrative Genome Viewer (IGV)[6]. While erroneously mapped mouse reads caused a number of falsely predicted SNPs in the bulk kidney cancer xenograft, these SNPs were completely missing after MCD. In addition, MCD also improved the prediction of high-impact SNPs. For example, mouse reads mapped to the human reference genome in the bulk tumor sample resulted in the wrongly predicted destruction of the start codon of the GRIA3 gene (fig. 5B).

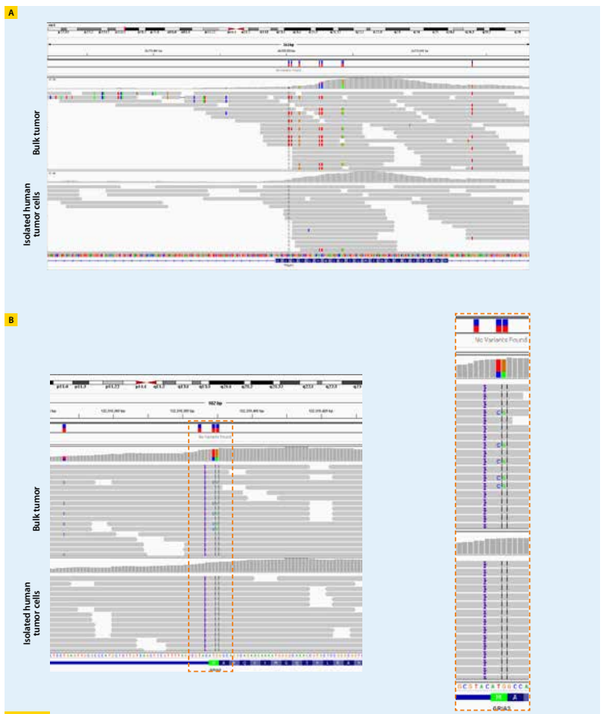

Figure 5

Conclusion

- We have developed an easy and fast (<20 min) cell separation method which allows for accurate downstream analysis of human tumor xenografts, avoiding bias caused by contaminating mouse cells.
- The contaminating mouse cells are specifically labeled prior to their depletion from the dissociated xenograft tissue. Labeling of the human cells is not required. Therefore, the procedure can be used for the isolation of all kinds of xenografted human material without the need for a positive marker expressed on the human cells.
- Removal of mouse cells significantly improves the analysis of human tumor xenografts by next-generation sequencing. As this effect was observed although a targeted human sequence–specific selection has been carried out during exome enrichment, the influence on whole exome and whole transcriptome sequencing are expected to be even more prominent.

References
1. DeRose, Y.S. *et al.* (2011) Nat. Med. 17: 1514–1520.
2. Bolger, A.M. *et al.* (2014) Bioinformatics 30: 2114–2120.
3. Li, H. and Durbin, R. (2009) Bioinformatics 25: 1754–1760.
4. http://broadinstitute.github.io/picard/
5. Koboldt, D.C. *et al.* (2013) Curr. Protoc. Bioinformatics 44: 15.4.1–15.4.17.
6. Robinson, J.T. *et al.* (2011) Nat. Biotechnol. 29: 24–26

Ministerium für Innovation, Wissenschaft und Forschung des Landes Nordrhein-Westfalen
EUROPEAN UNION
EUROPEAN REGIONAL DEVELOPMENT FUND
INVESTING IN YOUR FUTURE