Supplemental results and discussion

To estimate the level of similarity between CTC lines and their tumour of origin, we analysed in detail the RNAseq data of CTC44 and that of the cell line derived from the primary colon tumor of the same patient (CPP44). Data from all three replicates for each cell line was analysed independently to ensure robustness of the analysis (Supplementary table 5). 5,808,372 bases corresponded to similar regions in both cell lines and passed our filtering criteria for coverage quality (cf supplementary Methods). 99.85% of these bases were homozygous wild-type alleles present in both CTC44 and CPP44. 0.106% (6147 bases) represented either homozygous variants present in CTC44 and CPP44 or heterozygous alleles with the same allele representation in both cell lines. 0.020% (1175 bases) were homozygous or heterozygous variants from CPP44 that exhibited reduced frequency in CTC44. Finally, only a very small number of bases (1319, 0.023%) had conflicting alleles between CPP44 and CTC44. Thus, 15% of all detected variants were discordant between the CTC and primary tumor-derived cells. This is in line with previously reported proportion of discordant mutations detected between primary tumors and CTCs in other studies (eg 14% discordance in ERBB2 mutations in [1] and falls below the rate of discordant variants between matching primary and metastatic colorectal tumors identified using NGS (eg. 21% discordance for non-synonymous somatic mutations and indels in [2], and 22% discordance in [3]. Although the metastasis sample was not available for this patient, the high concordance between cDNA sequence from CTC44 and CPP44 suggests that CTCs isolated from this patient may have emerged from the primary tumor rather than from a metastasis.

It is of note that detected variants included a rare TP53 mutation (R273H) that is heterozygous in CPP44 but homozygous in CTC44. The presence in both primary and CTC cells of this variant, typically very rare in colorectal cancer, strongly corroborates the biological link between CTC44 and the cell line derived from its tumor of origin. In contrast, another rare variant (KRAS G12V) was detected as the slightly dominant (frequency 0.55-0.65) but heterozygous variant in CPP44 but was absent in CTC44. Since CPP44 and CTC44 are grown under similar conditions it strongly suggests that the absence of cells carrying this variant is unlikely to result from culture conditions. Altogether the restricted pattern of these variants in the CTC line suggests that CTCs circulating in this patient only represented a fraction of cells from the primary tumor, which has already been reported by others (*[4]*).

Taken together these findings indicate that the isolation process of CTCs described in the present study does not alter the genomic profile of these cells compared to matching primary tumors and is unlikely to induce strong clonal selection among these cells.

1. Polzer, B., et al., *Molecular profiling of single circulating tumor cells with diagnostic intention.* EMBO Mol Med, 2014. **6**(11): p. 1371-86.
2. Brannon, A.R., et al., *Comparative sequencing analysis reveals high genomic concordance between matched primary and metastatic colorectal cancer lesions.* Genome Biol, 2014. **15**(8): p. 454.
3. Vignot, S., et al., *Comparative analysis of primary tumour and matched metastases in colorectal cancer patients: evaluation of concordance between genomic and transcriptional profiles.* Eur J Cancer, 2015. **51**(7): p. 791-9.

4. Krebs, M.G., et al., *Molecular analysis of circulating tumour cells-biology and biomarkers.* Nat Rev Clin Oncol, 2014. **11**(3): p. 129-44.