Applications of genetic-epigenetic tissue mapping for plasma DNA in prenatal testing, transplantation and oncology

Wanxia Gai[1,2,3], Ze Zhou[1,2], Sean Agbor-Enoh[4,5,6], Xiaodan Fan[7], Sheng Lian[7], Peiyong Jiang[1,2,3], Suk Hang Cheng[1,2], John Wong[8], Stephen L Chan[9], Moon Kyoo Jang[4,6], Yanqin Yang[4,6], Raymond HS Liang[10], Wai Kong Chan[11], Edmond SK Ma[11], Tak Y Leung[12], Rossa WK Chiu[1,2], Hannah Valantine[4,6], KC Allen Chan[1,2,3], YM Dennis Lo[1,2,3]*

[1]Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong, China; [2]Department of Chemical Pathology, The Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong, China; [3]State Key Laboratory of Translational Oncology, The Chinese University of Hong Kong, Hong Kong, China; [4]Genomic Research Alliance for Transplantation (GRAfT), Bethesda, United States; [5]Division of Pulmonary and Critical Care Medicine, The Johns Hopkins School of Medicine, Baltimore, United States; [6]Division of Intramural Research, National Heart, Lung and Blood Institute, Bethesda, United States; [7]Department of Statistics, The Chinese University of Hong Kong, Hong Kong, China; [8]Department of Surgery, The Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong, China; [9]Department of Clinical Oncology, The Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong, China; [10]Comprehensive Oncology Centre, Hong Kong Sanatorium & Hospital, Hong Kong, China; [11]Department of Pathology, Hong Kong Sanatorium & Hospital, Hong Kong, China; [12]Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong, China

*For correspondence: loym@cuhk.edu.hk

Abstract We developed genetic-epigenetic tissue mapping (GETMap) to determine the tissue composition of plasma DNA carrying genetic variants not present in the constitutional genome through comparing their methylation profiles with relevant tissues. We validated this approach by showing that, in pregnant women, circulating DNA carrying fetal-specific alleles was entirely placenta-derived. In lung transplant recipients, we showed that, at 72 hr after transplantation, the lung contributed only a median of 17% to the plasma DNA carrying donor-specific alleles, and hematopoietic cells contributed a median of 78%. In hepatocellular cancer patients, the liver was identified as the predominant source of plasma DNA carrying tumor-specific mutations. In a pregnant woman with lymphoma, plasma DNA molecules carrying cancer mutations and fetal-specific alleles were accurately shown to be derived from the lymphocytes and placenta, respectively. Analysis of tissue origin for plasma DNA carrying genetic variants is potentially useful for noninvasive prenatal testing, transplantation monitoring, and cancer screening.

Introduction

The circulation receives DNA from different tissues and organs within the body. The analysis of plasma DNA from specific tissues or organs is useful for revealing and monitoring the pathological processes in different tissues. In scenarios where the genetic composition of a target tissue or organ

is different from the host constitutional genome, plasma DNA carrying the tissue- or organ-specific variants can be used to identify DNA molecules released by the tissue or organ. For example, in pregnant women, plasma DNA carrying fetal-specific alleles can be used for prenatal analysis of the fetal genetic constitution (*Kitzman et al., 2012*; *Lo et al., 2010*). In organ transplant recipients, the concentrations of donor-specific DNA has been used to reflect the tissue damage associated with acute rejection (*De Vlaminck et al., 2014*; *De Vlaminck et al., 2015*; *Knight et al., 2019*; *Lo et al., 1998*; *Schütz et al., 2017*). Notably, immediately after organ transplantation, the plasma concentration of donor-derived DNA surges (*De Vlaminck et al., 2015*). Because of this initial surge, the analysis for donor-derived DNA has limited value in identifying graft rejection and infection during the first 60 days of transplantation (*De Vlaminck et al., 2015*). The exact mechanism of this initial surge is unclear. It is possible that the hematopoietic cells within the transplanted organ are more likely to release a significant amount of DNA into the circulation during the initial days after the transplantation. However, existing methods for detecting DNA derived from a transplanted organ in plasma rely on identifying genetic differences between the organ donor and the recipient (*De Vlaminck et al., 2014*; *De Vlaminck et al., 2015*; *Knight et al., 2019*; *Lo et al., 1998*; *Schütz et al., 2017*). These methods cannot be used to further distinguish the exact cell types the donor DNA is derived from.

In situations where the genetic compositions of the different organs are the same, tissue composition analysis based on detecting organ-specific alleles would not be applicable. To overcome this, recent efforts have been made to measure the composition of DNA using epigenetic approaches. These approaches include methylation deconvolution (*Moss et al., 2018*; *Sun et al., 2015*), mapping nucleosomal patterns (*Snyder et al., 2016*; *Sun et al., 2019*), analysis of end DNA motifs, end positions and jaggedness (*Chan et al., 2016*; *Jiang et al., 2018*; *Jiang et al., 2020b*; *Jiang et al., 2020a*), and the profiling of RNA transcripts (*Koh et al., 2014*; *Tsui et al., 2014*). In these methods, the features of interest, for example, methylation patterns, of the plasma DNA were profiled and compared with those of the candidate tissues. Then the relative contribution of the different tissues to the circulating DNA was determined mathematically. One potential application of plasma DNA tissue composition analysis is to reveal the likely location of a concealed cancer. Recently, it has been shown that the analysis for circulating cell-free tumor DNA (ctDNA) is useful for the screening of early asymptomatic cancers (*Chan et al., 2017*; *Lennon et al., 2020*; *CCGA Consortium et al., 2020*). As cancer-associated genetic and epigenetic changes are present in virtually all types of cancers (*Chan et al., 2013a*; *Chan et al., 2013b*; *Leary et al., 2012*; *Wong et al., 1999*), the detection of these cancer-associated aberrations in plasma can potentially serve as universal tumor markers for the screening of cancers in general. However, how subjects with positive results of a universal cancer test can be further worked up is an important but relatively under-explored topic. In a study by Lennon et al., subjects tested positive with ctDNA test that detected a wide variety of cancers were investigated with whole body positron emission tomography-computed tomography (PET-CT) (*Lennon et al., 2020*). If the potential tissue origin of the cancer can be obtained from ctDNA analysis, more focused investigations, for example, high-resolution imaging of an affected organ, can be performed. These organ-specific investigations could provide better sensitivity and specificity and could be achieved with a lower dose of radiation to the patients. In previous proof-of-principle studies, the tissue origin of cancers was successfully revealed by plasma DNA deconvolution (*CCGA Consortium et al., 2020*; *Moss et al., 2018*; *Sun et al., 2015*). However, existing approaches only allow tissue composition analysis of the whole pool circulating DNA rather than specifically to the tumor-derived DNA. The accuracy of these approaches would be affected by the fractional concentration of tumor-derived DNA in the sample.

In this study, we developed a method called genetic-epigenetic tissue mapping (GETMap) to determine the tissue composition of plasma DNA carrying genetic variants which are different from the host constitutional genome. This method is based on the comparison of the methylation profiles of the plasma DNA carrying genetic variants and the relevant tissues or organs that plasma DNA is potentially derived from. First, we validated this approach using a pregnancy model through the analysis of the tissue origin of the plasma DNA carrying fetal-specific alleles. Then, we applied this method to measure the tissue compositions of plasma DNA carrying cancer-associated mutations (i.e., present in tumor cells or plasma but absent from buffy coats) in hepatocellular cancer (HCC) patients and those molecules carrying donor-specific alleles in lung transplant recipients. The former analysis can provide information regarding the tissue origin of the cancer and the latter analysis

provided insights on the reason for the surge of donor-derived DNA in the plasma of organ transplant recipients during the early post-transplantation period.

Results

Principle of GETMap

The principle of the GETMap analysis is illustrated in *Figure 1*. The first step is to identify different sets of plasma DNA molecules based on genotypic differences. For example, the two sets of plasma DNA molecules carrying cancer-associated mutations and wildtype alleles were identified in cancer patients. In organ transplant recipients, three sets of DNA molecules can be identified, including those carrying the host-specific, recipient-specific alleles and alleles shared between the host and recipient. Similarly, three sets of molecules could be identified in the plasma of a pregnant woman, namely those carrying fetal-specific, maternal-specific alleles and alleles shared by the mother and fetus. Then, the tissue compositions were determined for each set of plasma DNA molecules through comparing the methylation profile of the plasma DNA molecules and the methylation profiles of the relevant tissues after bisulfite sequencing. While there are some similarities between the deconvolution step and that described in our previous study (*Sun et al., 2015*), there are notable differences. First, only DNA molecules of interest, for example, those carrying fetal-specific alleles, or cancer-associated mutations or donor-derived alleles, are analyzed. Second, only CpG sites near informative single nucleotide polymorphism (SNP) alleles are included in the algorithm. The details of the mathematical calculation are described in the 'Materials and methods' section. For the choice of candidate tissues used for the GETMap analysis, we included the tissues (including neutrophils, lymphocytes, liver, and placenta) that have been validated in a previous study on tissue deconvolution by methylation analysis (*Sun et al., 2015*). The inclusion of the placenta also allows us to use the analysis of fetal DNA in maternal plasma as a model to validate this new approach. As this study also analyzed patients receiving lung transplantation, lung is further included as one candidate tissue in the plasma DNA deconvolution. The methylation status of the plasma DNA molecules was determined by bisulfite sequencing.

Accuracy of GETMap analysis

To evaluate the accuracy of our approach, we performed simulation analyses using GETMap to deconvolute five types of reference tissues including neutrophils, lymphocytes, lung, liver, and placenta. Three sets of simulation analyses were performed to simulate the three clinical application scenarios in our study, namely pregnancy, transplantation, and cancer detection. For each scenario, the numbers of informative DNA fragments, CpG sites, and sequencing depth were matched with the median of the studied samples. Thirty independent simulations were performed for each scenario. The accuracy was calculated as the percentage contribution assigned to the tissue used for the deconvolution. For example, when the bisulfite sequencing data of liver tissue is used for deconvolution, the accuracy would refer to the estimated contribution from liver. The median accuracy of GETMap analyses for reference tissues was 98.3% (range 95.5–99.8%) (*Table 1*).

Deconvolution of fetal- and maternal-derived DNA in maternal plasma

We first used the analysis of plasma DNA of pregnant women as a model to demonstrate the feasibility of GETMap. Venous blood samples were collected from 30 pregnant women with 10 in each of the first, second, or third trimesters of gestation. Placental tissues were obtained from chorionic villus sampling or amniocentesis for the first and second trimester pregnant women. For third trimester pregnant women, the placenta was collected after delivery. The pregnant woman and the placental tissue were genotyped using the Illumina whole-genome arrays (HumanOmni2.5, Illumina). Based on the genotypes of the mother and fetus, we identified a median of 189,862 (range 14,035–192,998) maternal-specific informative SNPs where the mother was heterozygous and the fetus was homozygous, and a median of 194,479 (range 145,743–201,847) fetal-specific informative SNPs where the mother was homozygous and the fetus was heterozygous. After bisulfite sequencing of maternal plasma DNA, a median of 103 million uniquely mapped reads (range: 52–186 million) were identified in the maternal plasma DNA samples. Plasma DNA molecules carrying the fetal- and maternal-specific alleles were identified. A median of 162,813 CpG sites (range 8237–295,671) and 53,039 CpG
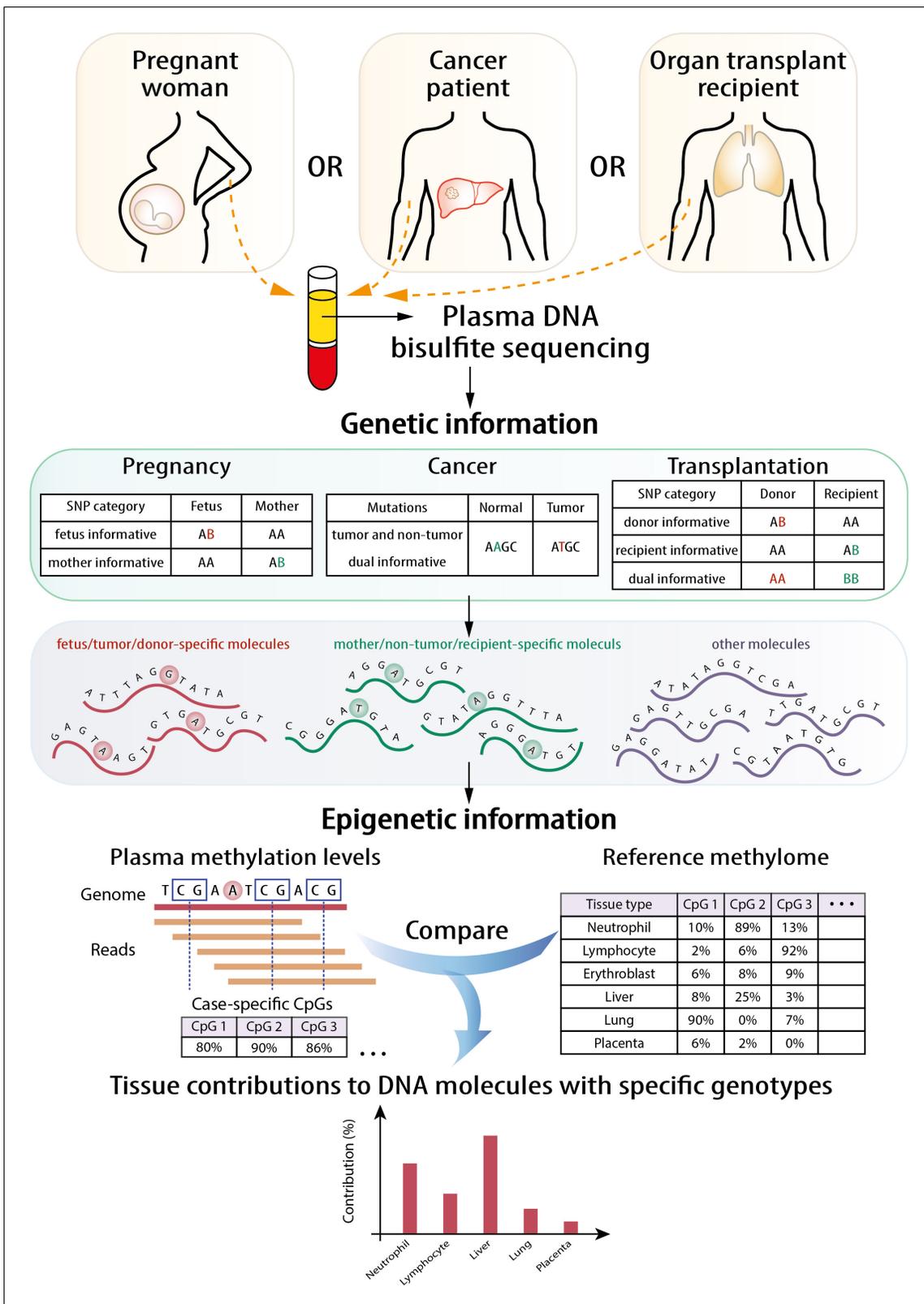
Figure 1. Schematic illustration of the principle of genetic-epigenetic tissue mapping (GETMap) analysis. The paired individuals (e.g., fetus/mother, organ donor/recipient, and tumor/normal tissue) are genotyped to identify single nucleotide polymorphism (SNP) alleles specific for one of them. After bisulfite sequencing, plasma DNA molecules carrying individual-specific alleles and at least one CpG site are identified. The plasma DNA methylome is

Figure 1 continued on next page

Figure 1 continued

compared with the methylation profiles of reference tissues to determine the tissue composition of the subset of plasma DNA molecules derived from a particular individual.

sites (range 16,796–138,284) were identified on the plasma DNA molecules carrying maternal-specific and fetal-specific alleles, respectively. For the plasma DNA molecules carrying fetal-specific alleles, the median deduced contribution from the placenta was 100% (*Figure 2A*). These results are compatible to the results of previous studies that fetal DNA in maternal plasma is derived from the placenta (*Alberry et al., 2007*; *Masuzaki et al., 2004*). For molecules carrying maternal-specific alleles, a median of 80% of DNA molecules were deduced to be derived from hematopoietic cells (i. e., neutrophils and lymphocytes) (*Figure 2B*). All cases showed no contribution from the placenta. For molecules carrying the shared alleles at SNPs where the mother was homozygous and the fetus was heterozygous, the deduced placental contribution showed a positive correlation with the fetal DNA fractions based on the ratio between the number of plasma DNA molecules carrying fetal-specific alleles and alleles shared by the mother and the fetus (*Figure 2C*).

Deconvolution of DNA molecules carrying donor- and recipient-specific alleles following lung transplantation

We applied GETMap analysis to patients who had received lung transplantation and explored if the tissue composition would change over time. Forty samples from 11 patients were collected (*Table 2*). By comparing the SNP genotypes between the donor and recipient, we identified a median of 270,144 (range 254,846–344,024) donor-specific informative SNPs where the donor was heterozygous and the recipient was homozygous and a median of 270,285 (range 261,529–357,009)

Table 1. Results of deconvolution of bisulfite sequencing data from reference tissues for scenarios of (A) pregnancy, (B) lung transplantation, and (C) liver cancer.

The underlined numbers represent the percentage of contribution accurately assigned to the respective tissues by genetic-epigenetic tissue mapping (GETMap).

(A)

		Tissue contribution as determined by GETMap analysis				
		Neutrophils	Lymphocytes	Liver	Lung	Placenta
Reference tissue used for the simulation	Neutrophils	96.78	2.01	0.59	0.33	0.29
	Lymphocytes	0.52	98.30	0.41	0.20	0.58
	Liver	0.31	0.64	98.36	0.27	0.42
	Lung	0.24	0.66	0.35	98.36	0.39
	Placenta	0.13	0.05	0.00	0.09	99.73

(B)

		Tissue contribution as determined by GETMap analysis				
		Neutrophils	Lymphocytes	Liver	Lung	Placenta
Reference tissue used for the simulation	Neutrophils	98.21	0.77	0.42	0.43	0.17
	Lymphocytes	0.48	98.70	0.20	0.31	0.31
	Liver	0.32	0.19	99.25	0.11	0.13
	Lung	0.21	0.09	0.22	99.39	0.09
	Placenta	0.00	0.09	0.08	0.05	99.78

(C)

		Tissue contribution as determined by GETMap analysis				
		Neutrophils	Lymphocytes	Liver	Lung	Placenta
Reference tissue used for the simulation	Neutrophils	96.08	2.23	0.32	0.37	1.00
	Lymphocytes	0.94	95.46	0.79	2.06	0.75
	Liver	0.50	0.44	96.67	1.48	0.91
	Lung	0.90	1.71	0.80	96.08	0.51
	Placenta	0.49	0.13	0.77	0.34	98.27
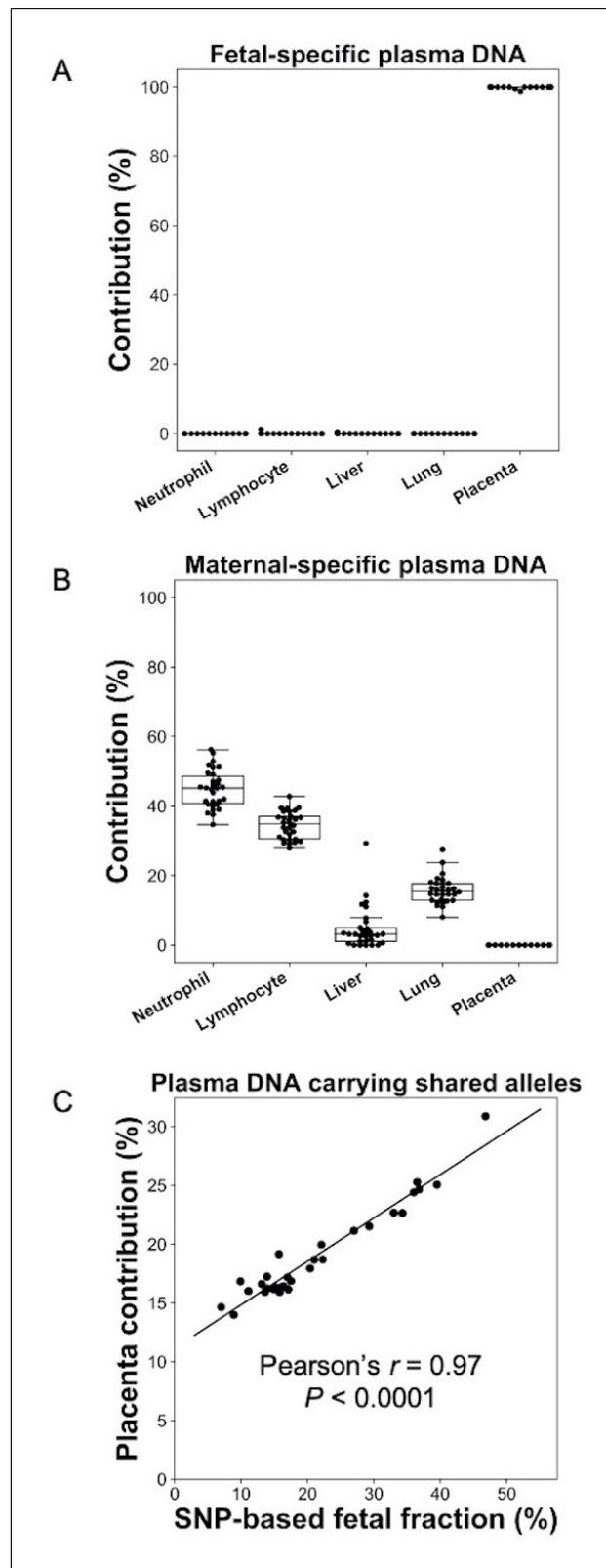
Figure 2. Percentage contributions of different cell types to maternal plasma DNA carrying (A) fetal-specific alleles and (B) maternal-specific alleles in 30 pregnant women. (C) Correlation between percentage contribution of the placenta to maternal plasma DNA molecules carrying alleles shared by the fetus and mother and single nucleotide polymorphism (SNP)-based fetal DNA fraction.

Table 2. The demographic profiles of lung transplant recipients.

Case number	Recipient age	Recipient gender	Donor age	Donor gender	Diagnosis for transplant	Single/ double lung	Cause of death	Time of sample collection post-transplant
1	34	M	32	M	Cystic fibrosis	Double	Alive	72 hr
2	59	F	27	F	Interstitial lung disease	Double	Alive	72 hr
3	53	M	20	M	Interstitial lung disease	Double	Alive	72 hr
4	63	M	16	F	Interstitial lung disease	Double	Alive	72 hr, 6 dy
5	55	F	36	F	Interstitial lung disease	Double	Alive	72 hr, 7 dy
6	66	M	48	F	Interstitial lung disease	Single	Alive	72 hr, 4 wk
7	66	F	18	M	Chronic obstructive pulmonary disease	Single	Alive	72 hr, 7 dy, 5 wk, 20 wk, 25 wk, 157 wk
8	32	F	39	M	Cystic fibrosis	Double	Alive	72 hr, 7 dy, 8 wk, 38 wk, 77 wk, 129 wk
9	67	F	53	F	Sarcoidosis	Double	Respiratory failure	72 hr, 7 dy, 6 wk, 13 wk, 22 wk
10	44	M	35	F	Retransplant	Double	Alive	72 hr, 7 dy, 10 dy, 4 wk, 14 wk, 25 wk, 103 wk
11	67	F	32	M	Pulmonary arterial hypertension	Single	Alive	72 hr, 7 dy, 5 wk, 15 wk, 26 wk, 61 wk, 104 wk

*Samples collected when the patient was having a rejection episode were underlined.

recipient-specific informative SNPs where the donor was homozygous and the recipient was heterozygous. In addition, a median of 81,957 (range 77,196–133,422) dual informative SNPs where both the donor and recipient were homozygous but for different alleles were identified. After bisulfite sequencing of the plasma DNA, a median of 327 million uniquely mapped reads (range 32–481 million) were obtained for each case. A median of 920,830 (range 141,065–1,329,292) and 141,794 (range 12,700–529,211) CpG sites were identified on the plasma DNA molecules carrying recipient- and donor-specific alleles, respectively.

For each subject, the first sample was collected at 72 hr after the transplantation. We performed the GETMap analysis on donor-derived DNA molecules for each sample collected at 72 hr post-transplant (*Figure 3A*). The median contribution from the lung to the donor-derived DNA was only 17%. Surprisingly, a substantial proportion of the DNA molecules carrying the donor-specific alleles were contributed from the hematopoietic cells. The median contribution from the neutrophils and lymphocytes combined was 78%. The median deduced contribution from all other tissues was 5% in total.

We studied the changes in the lung DNA proportions in the donor-derived plasma DNA molecules with time after transplantation. We categorized the samples based on the time of sample collection post-transplant: within 72 hr; in-between 72 hr, 7 days, 10 weeks, and 50 weeks; and beyond 50 weeks. The 40 samples were thus classified into five categories that included 11, 7, 7, 9, and 6 samples, respectively. The median fractional concentrations of donor-derived DNA were 16%, 6%, 2%, 1%, and 2% for these categories, respectively (*Figure 3B*). The median contributions from the lung to the donor-derived DNA were 17%, 34%, 59%, 51%, 66% for samples in these categories, respectively (*Figure 3C*). These data showed that the lung DNA proportions in donor-derived DNA increased with time after transplantation. In contrast, the median contributions from the hematopoietic cells decreased with time, that is, 78%, 56%, 27%, 41%, and 21% for samples in the five categories, respectively. For the plasma DNA molecules carrying the recipient-specific alleles, we observed the hematopoietic cells as the key contributors. For samples in the five categories, the median contributions of hematopoietic cells were 83%, 86%, 89%, 94%, and 84%, respectively (*Figure 4*).

We further explored if the fractional contribution of the lung to the donor-specific DNA would be useful for the detection of graft rejection. As all the rejection episodes occurred after 7 days, only samples collected after 7 days were used for this analysis. The median donor-derived DNA fractions were 3% for the samples collected during rejection episodes and 1% for those collected during remission (p-value=0.22, Mann-Whitney rank-sum test, *Figure 3B*). The median lung contributions
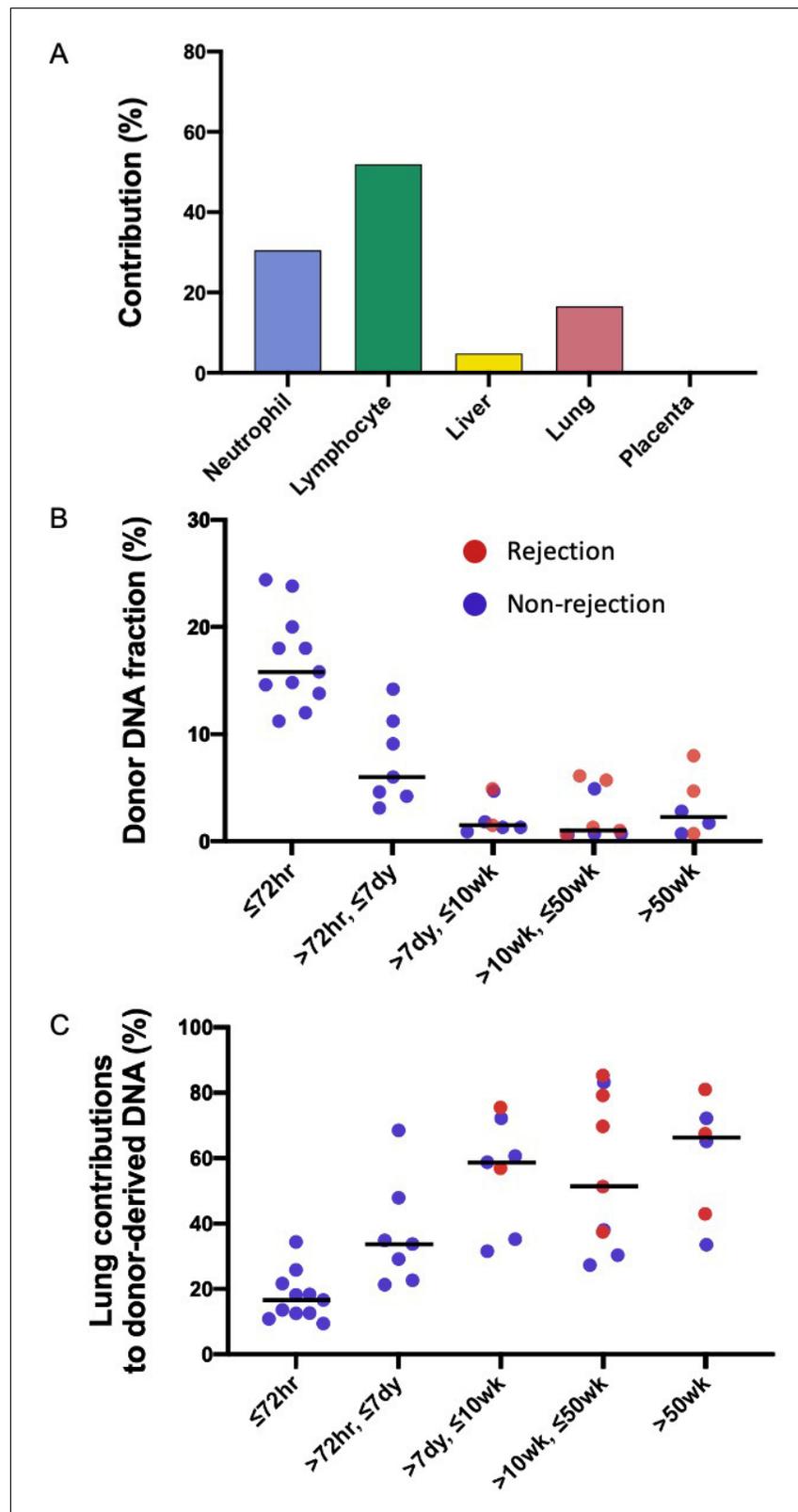
Figure 3. Genetic-epigenetic tissue mapping (GETMap) analysis on donor-derived plasma DNA molecules in lung-transplant recipients. (**A**) The median percentage contributions of different cell types to plasma DNA carrying donor-specific alleles in patients with lung transplantation at 72 hr post-transplant. (**B**) Fractional concentrations of
Figure 3 continued on next page

donor-derived DNA and (C) percentage contributions of the lung to plasma DNA carrying donor-specific alleles in patients with lung transplantation.

were 69% and 48% for these two groups of samples, respectively (p-value=0.09, Mann-Whitney rank-sum test, *Figure 3C*).

Deconvolution of plasma DNA molecules carrying mutant identified in tumor tissues

We then explored if GETMap analysis could reveal the tissue origin of ctDNA in two HCC patients. The two patients were denoted as HCC 1and HCC 2, respectively. In the initial analysis, we first identified the cancer-specific mutations by analyzing the tumor tissues and the buffy coat of the patients. A total of 30,383 and 6996 tumor-specific single nucleotide mutations were identified from HCC 1 and HCC 2, respectively. After bisulfite sequencing of plasma DNA, 245 and 188 million uniquely mapped reads were obtained for the two patients, respectively. The numbers of plasma DNA molecules carrying the mutant alleles were 29,868 and 5090, and these molecules covered 18,193 and 4076 CpG sites, respectively. Tissue contributions of these tumor-derived plasma DNA molecules were deduced by GETMap analysis (*Figure 5*). The liver was deduced to be the key contributor with 90% (HCC 1) and 87% (HCC 2). A small contribution of 10% (HCC 1) and 13% (HCC 2) was from the placenta. The numbers of molecules carrying the wildtype alleles were 153,238 and 26,792, containing 35,883 and 8156 CpG sites, respectively. The contribution of the hematopoietic cells was deduced to be 48% (HCC 1) and 53% (HCC 2) whereas the liver contributed 32% (HCC 1) and 23% (HCC 2).
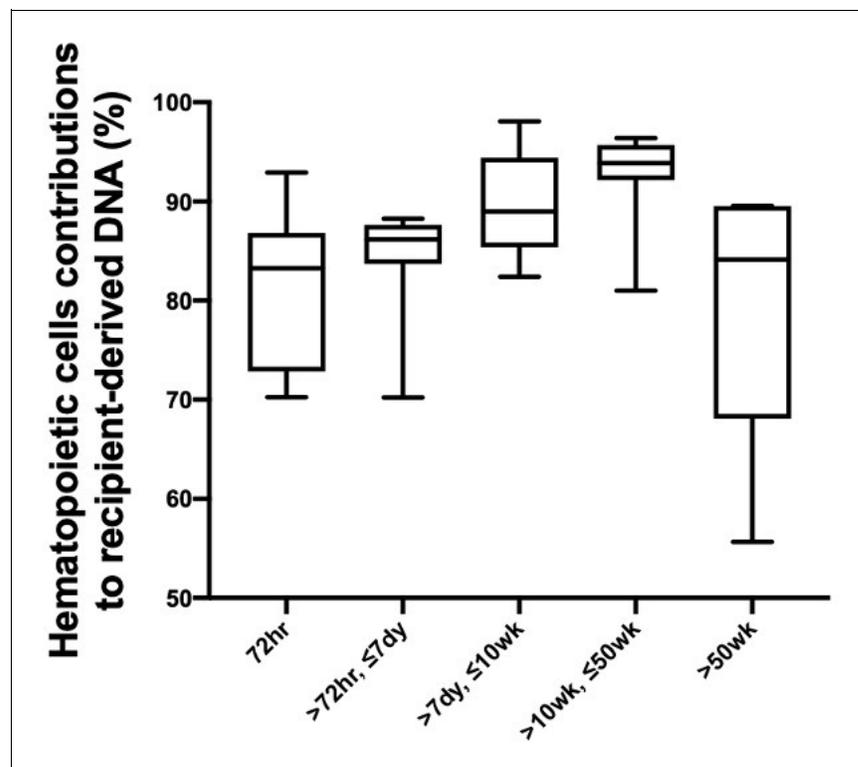
Figure 4. Percentage contributions of hematopoietic cells to the plasma DNA carrying recipient-specific alleles in patients with lung transplantation.
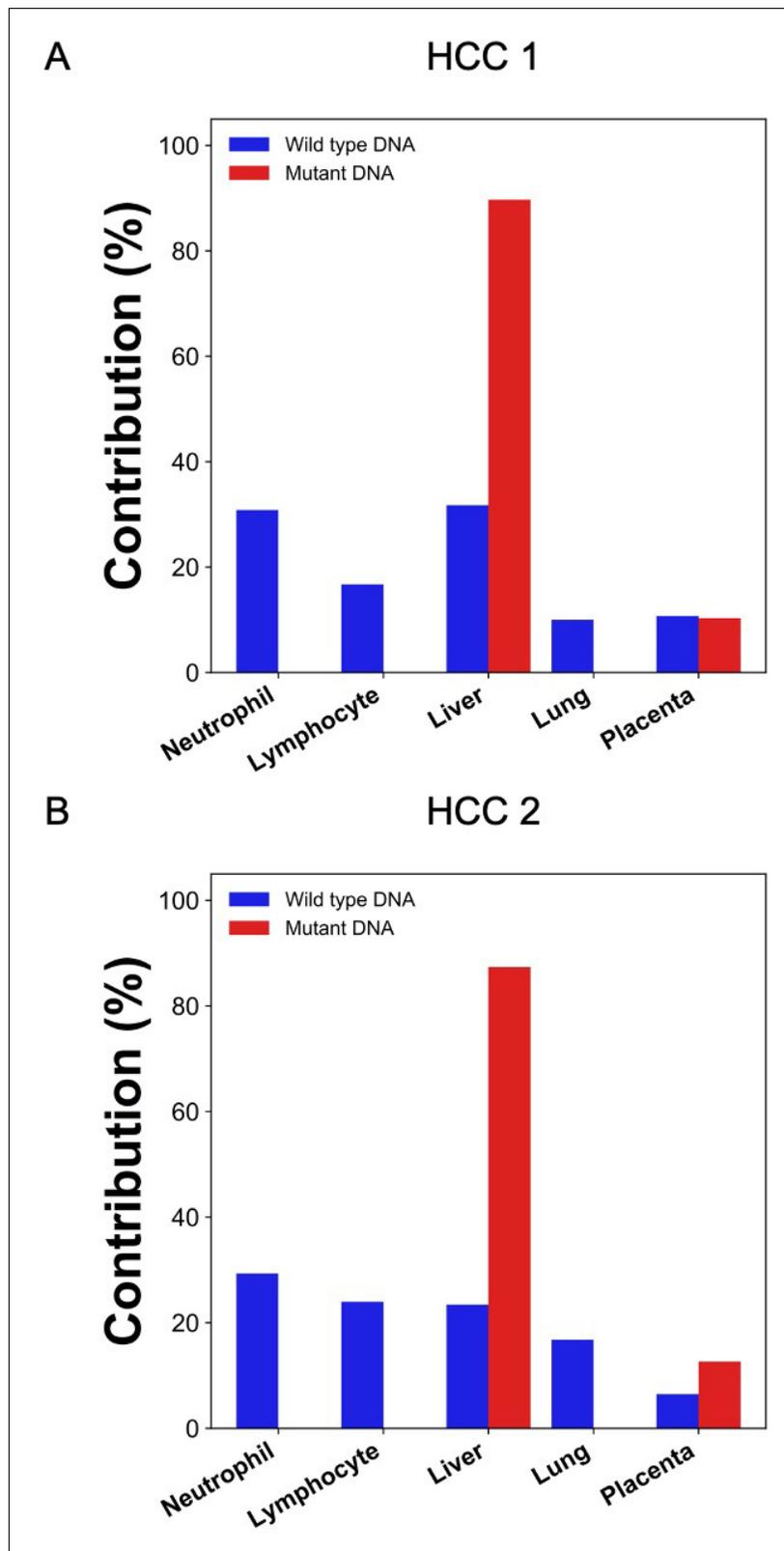
Figure 5. Percentage contributions of different tissues to plasma DNA with tumor-specific and wildtype alleles in two hepatocellular cancer (HCC) patients. The tumor-specific mutations were deduced from the tumor tissues.

Deconvolution of DNA carrying mutations directly derived from plasma

In the scenario of cancer screening using a universal tumor marker based on plasma DNA analysis, the tumor tissue would not be available for mutation analysis. Hence, we further explored if the cancer mutations can be directly derived from plasma DNA analysis. To obtain the mutation information directly from the plasma DNA, we sequenced the buffy coat and plasma DNA without bisulfite conversion. The sequencing depth for the plasma DNA were 50x and 61x haploid genome coverage and those for the buffy coat DNA were 53x and 55x in HCC 1 and HCC 2, respectively. Single nucleotides variations present in the plasma for more than a threshold number of occasions but not in the buffy coat were identified as candidate mutations (see details in the 'Materials and methods'). The numbers of candidate mutations identified were 10,864 and 3446 for the two HCC patients. GETMap analysis was then performed using the plasma DNA bisulfite sequencing data. The numbers of plasma DNA molecules carrying the cancer mutations were 16,200 and 4112, and covered 12,887 and 2991 CpG sites, respectively. For molecules carrying mutations, the contributions from the liver were estimated to be 69% (HCC 1) and 95% (HCC 2) (*Figure 6*). The placenta contributed the remaining proportion of 31% (HCC 1) and 5% (HCC 2). For molecules carrying wildtype alleles, hematopoietic cells, including neutrophils and lymphocytes, contributed a total of 51% (HCC 1) and 27% (HCC 2).

Deconvolution of plasma DNA for a pregnant woman with lymphoma

We previously reported the deconvolution results of total plasma DNA for a pregnant woman who was diagnosed as having follicular lymphoma during early pregnancy (*Sun et al., 2015*). In the current study, we explored if GETMap analysis could determine the tissue composition of the fetal- and cancer-derived DNA independently. We sequenced the lymphoma tissue, as well as the normal cells harvested from buccal swab and post-treatment buffy coat. As the pregnancy was terminated at time of the diagnosis of cancer, no placental tissue was collected. Hence, we deduced the fetal genotypes directly from the plasma DNA. Based on the non-bisulfite sequencing results of the plasma DNA and normal cells, 254,540 variants were identified in the plasma DNA. The algorithm for classifying these variants into fetal-specific alleles and cancer mutations is shown in *Figure 7*. We reasoned that variants overlapping with the common variations in the dbSNP Build 135 database were more likely derived from the fetus whereas those not overlapping with the database were more likely to come from the tumor. For the 13,546 variants that did not overlap with dbSNP database, 2641 were detected in three or more sequence reads of the tumor tissues. These variants are regarded as tumor mutations for GETMap analysis. For the 240,994 variants overlapping with the dbSNP database, 231,552 were completely absent in the tumor tissue. These variants were likely derived from the fetus and are regarded as fetal-specific alleles for the GETMap analysis. The allele frequencies for the fetal-specific SNPs and tumor-specific mutations in plasma were normally distributed and peaked at 6% and 20%, respectively (*Figure 8*).

After bisulfite sequencing of plasma DNA, we obtained 700 million uniquely mapped reads. We identified DNA molecules carrying the tumor-specific mutant alleles, wildtype alleles, fetal-specific alleles, and the alleles shared by the fetus and the mother. The GETMap analysis was performed on each set of plasma DNA molecules to deduce their tissue composition. The numbers of CpG sites covered by the DNA molecules carrying the mutant and wildtype alleles were 4781 and 6660, respectively. For the molecules carrying tumor mutations, it was deduced that 100% was from lymphocytes (*Figure 9A*). For molecules carrying the wildtype alleles, the deduced contribution from neutrophils, lymphocytes, liver, lung, and placenta were 29%, 46%, 13%, 2%, and 11%, respectively. For DNA molecules carrying the fetal-specific, the deduced contribution from the placenta was 95% (*Figure 9B*). For those carrying alleles shared by the mother and fetus, the deduced contribution from neutrophils, lymphocytes, liver, lung, and placenta were 23%, 48%, 11%, 14%, and 5%, respectively.

Discussion

In this study, we developed GETMap analysis to determine the tissue origin of plasma DNA molecules carrying genetic variants. In this method, we first identified a subset of plasma DNA molecules carrying specific alleles. Then, by comparing the methylation status of these molecules and the
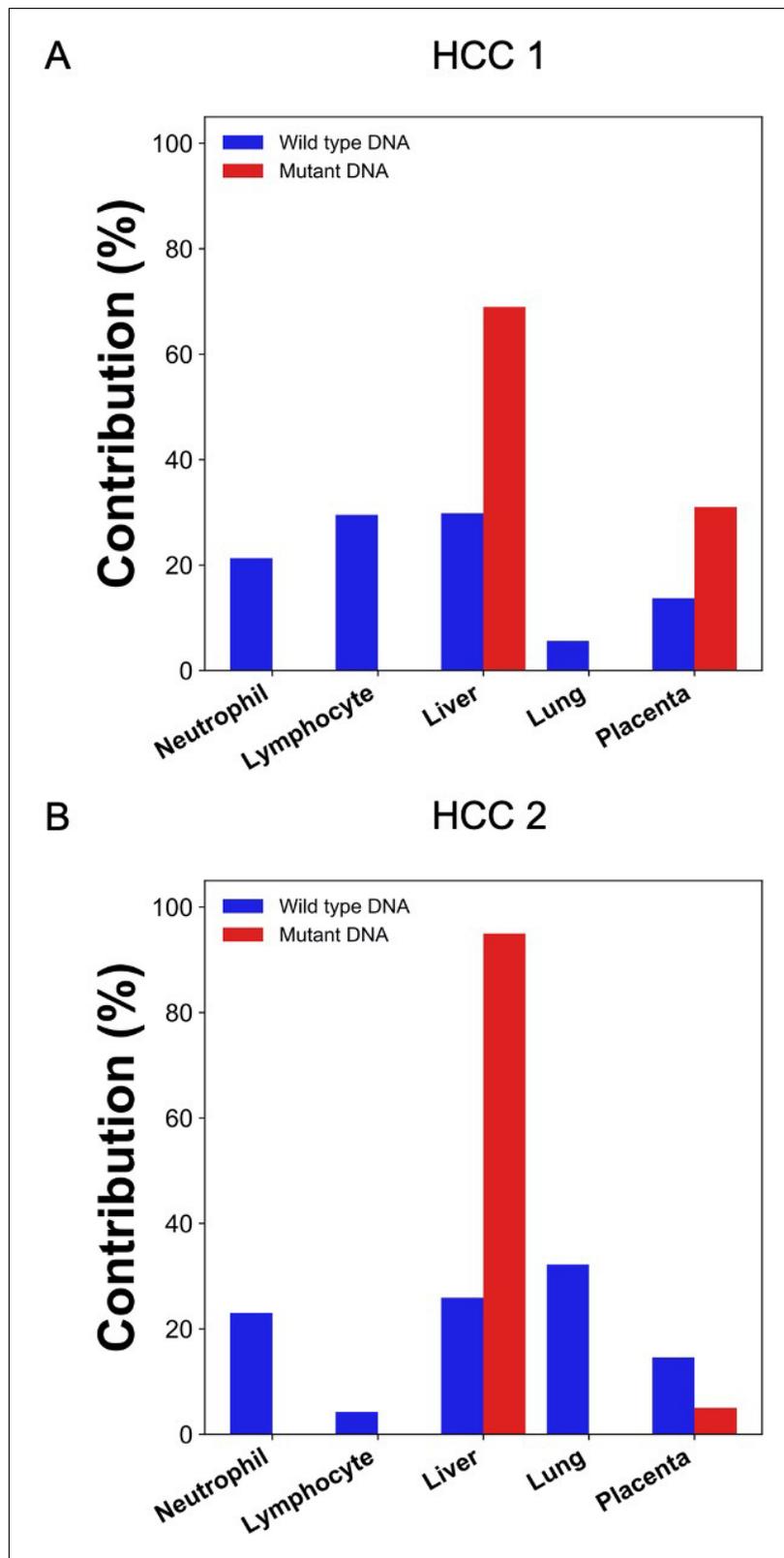
Figure 6. Percentage contributions of different tissues to plasma DNA with tumor-specific and wildtype alleles in two hepatocellular cancer (HCC) patients. The tumor-specific mutations were deduced directly from the plasma.
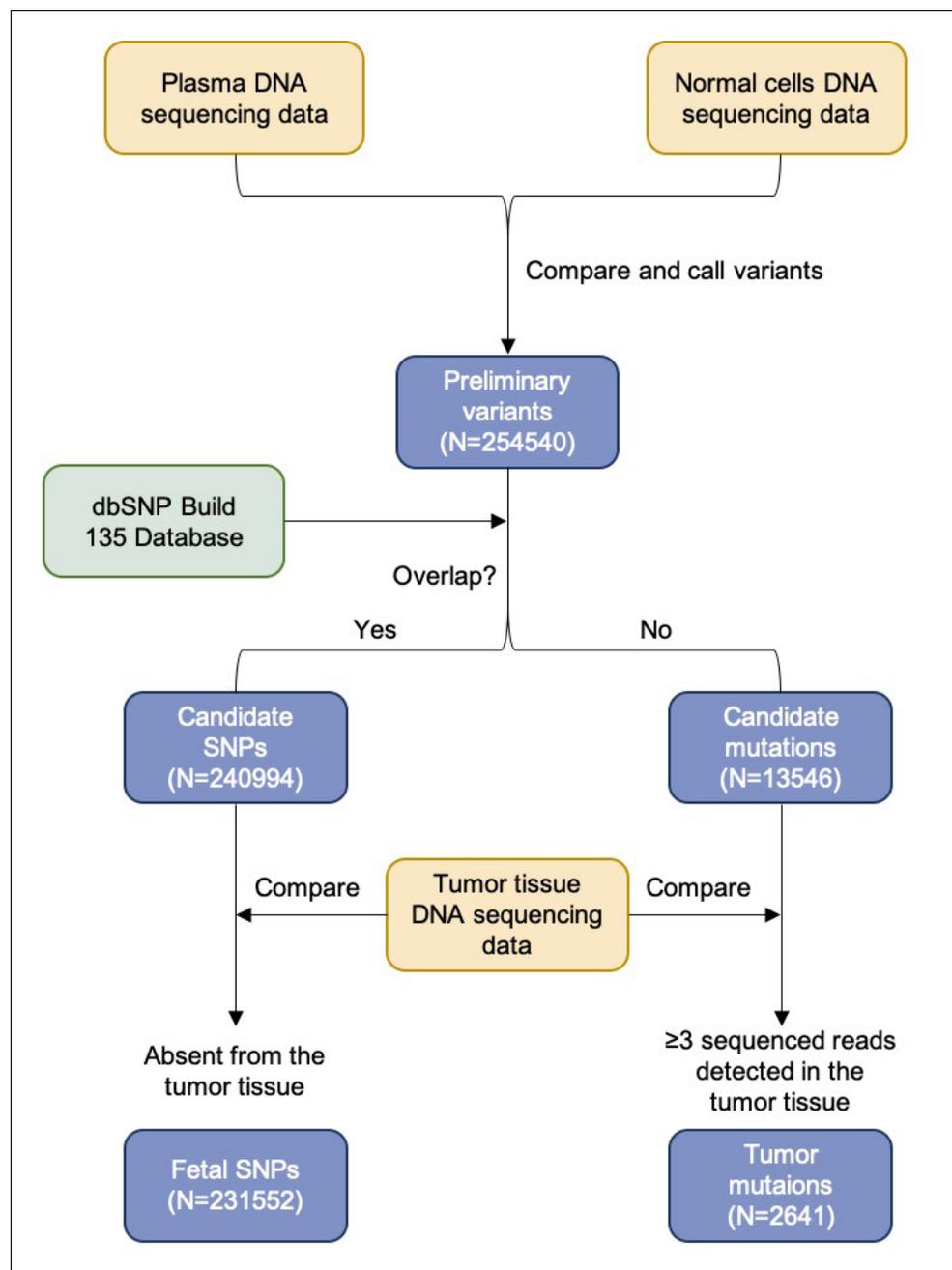
Figure 7. Flowchart of the steps for identifying the fetal-specific alleles and cancer mutations in the pregnant woman with lymphoma.

methylation profiles of the candidate tissue organs, we could determine the tissue composition of the DNA molecules. In the first part of the study, we used the pregnancy model to validate the GET-Map analysis. The plasma DNA molecules carrying the fetal-specific alleles were deduced to be 100% derived from the placenta. For the molecules carrying the alleles shared by the fetus and the mother, the percentage contribution from the placenta showed a positive linear relationship with the fractional concentration of fetal DNA based on SNP analysis. These results are consistent with the previous studies which showed that the fetal DNA in maternal plasma is indeed derived from the placenta. For the plasma DNA molecules carrying maternal-specific alleles, no contribution from the placenta was observed. A large proportion was derived from the hematopoietic cells, neutrophils, and lymphocytes, with a median total contribution of 80%. These figures are comparable to those
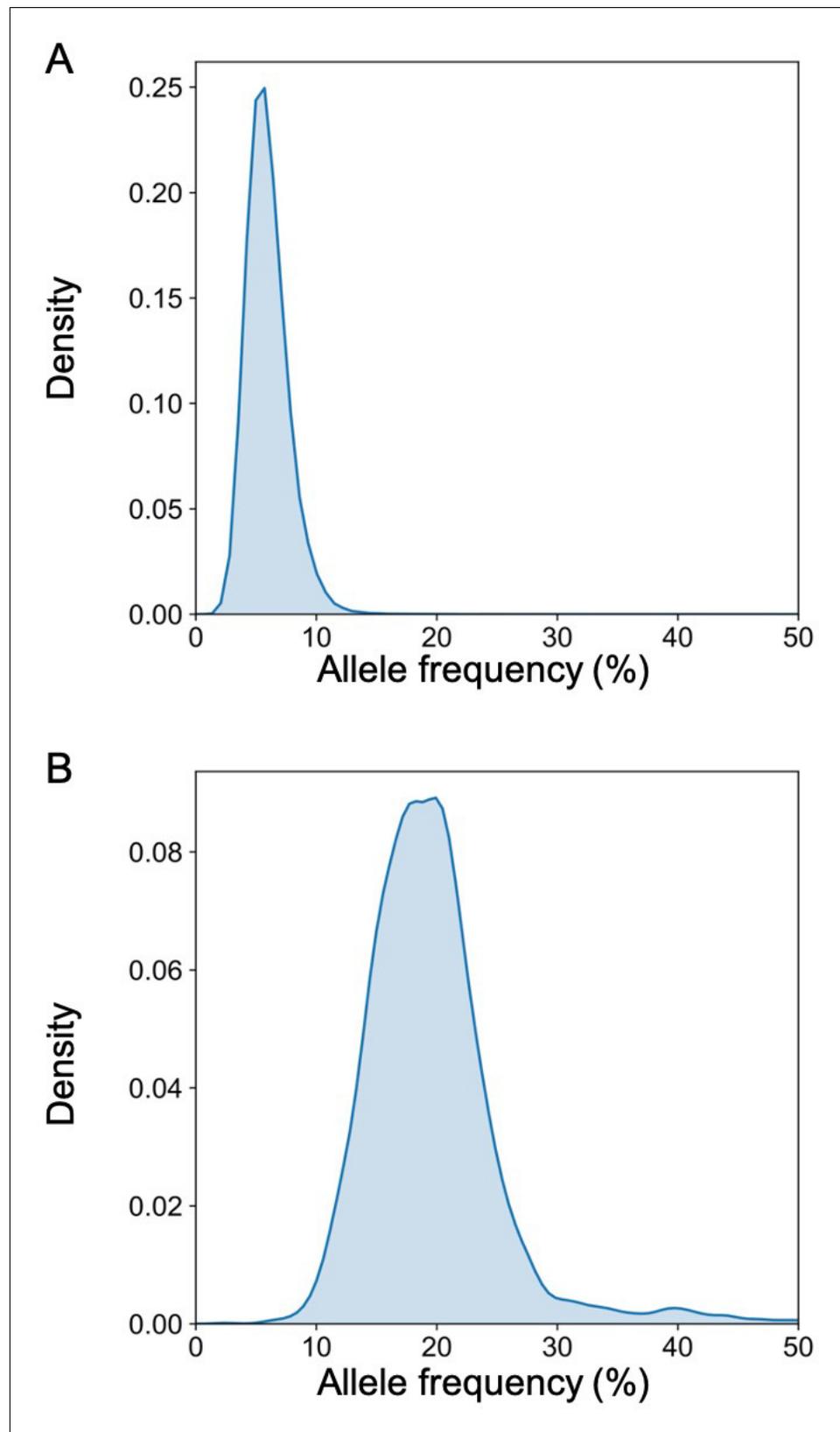
Figure 8. The distribution of the allele frequency of (A) the fetal-specific alleles and (B) the mutant alleles in the plasma of the pregnant woman with lymphoma.
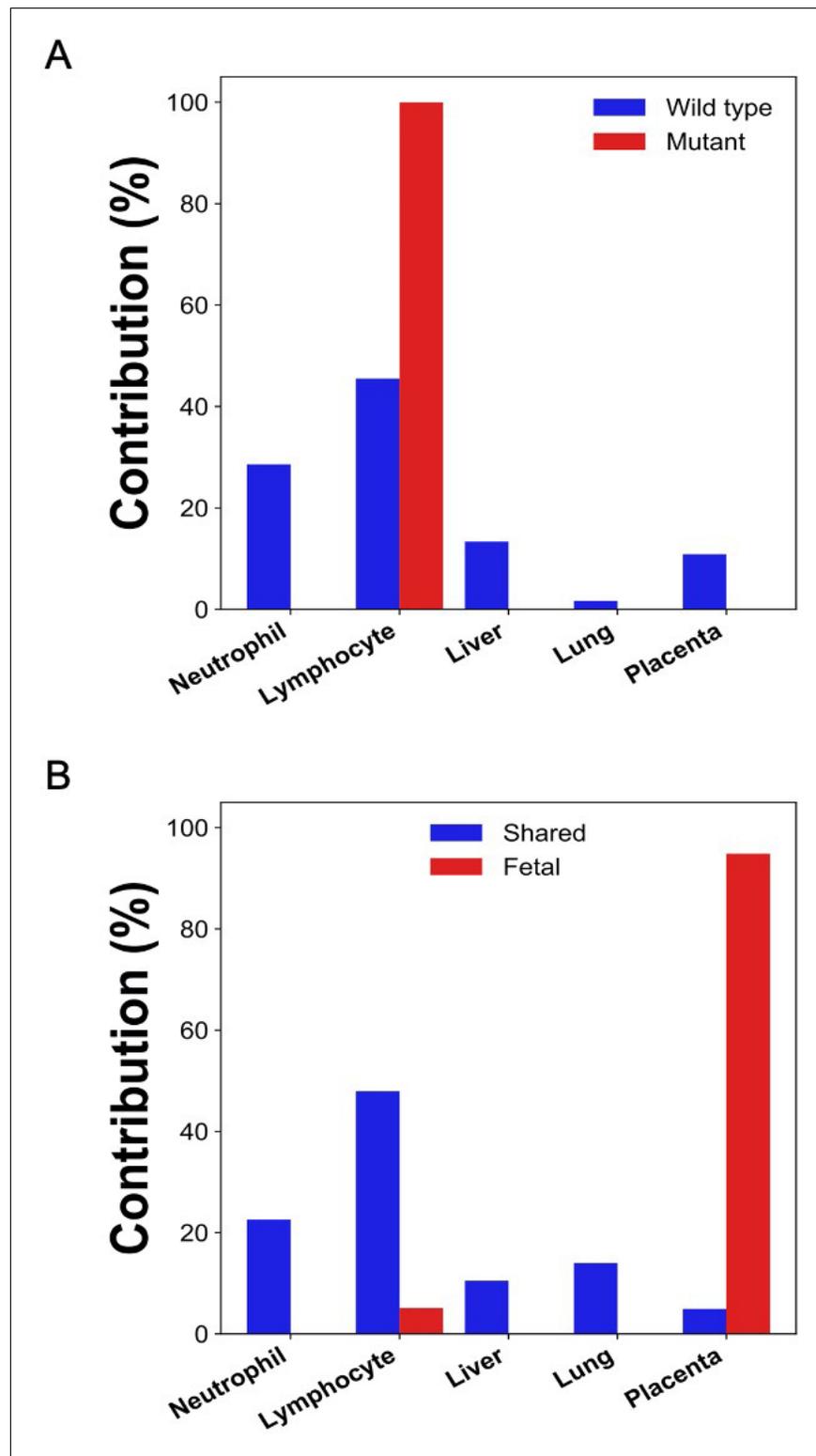
Figure 9. Percentage contributions of different tissues to (**A**) plasma DNA with tumor-specific and wildtype alleles, and (**B**) fetal-specific plasma DNA and DNA carrying the alleles shared by the fetus and the mother in a pregnant woman with lymphoma.

reported previously in healthy subjects (*Gai et al., 2018*; *Sun et al., 2015*). These results demonstrate the feasibility of determining the tissue contributions to the different genetic components of plasma DNA using GETMap analysis.

We then showed that, in patients who had received lung transplantation, a substantial proportion of donor-derived DNA was derived from the hematopoietic cells during the early post-transplant period. Previous studies have shown that a high level of DNA carrying donor genotypes would be present in the plasma of organ transplant recipients during the early post-transplant period even in the absence of any evidence of organ rejection (*De Vlaminck et al., 2015*). Hence, quantitative analysis for donor DNA in plasma cannot be used for reflecting transplant organ damage or rejection within 60 days of transplantation. The reason for this elevation in donor DNA was unclear. Using GETMap analysis, we determined the tissue composition of plasma DNA molecules carrying donor-specific alleles for samples collected at different time intervals after transplantation. Importantly, at 72 hr after transplantation, the median contribution from the lung was only 17% and a substantial contribution of 78% was from hematopoietic cells. This is likely due to the presence of residual blood cells in the transplanted organ and they could release DNA with donor genotypes into the circulation. The contribution of the lung gradually increases with time together with a parallel decline in the contribution of the hematopoietic cells. The median contribution of hematopoietic cells dropped to 21% after 50 weeks. The persistent contribution from the hematopoietic cells may be due to imprecision of measurement as the concentrations of donor DNA after 50 weeks were very low in patients without evidence of rejection. Alternatively, there may be persistence of donor hematopoietic cells in the body of the transplant recipient. In this regard, it has been shown that some immune cells resident in the donor tissue can be long-lived and self-renewing (*Gasteiger et al., 2015*). The lung fraction appeared to be higher for samples collected during graft rejection compared with those collected during remission. However, the difference did not reach statistical significance. Future studies with larger sample size would be useful to further explore this point.

We then investigated if GETMap analysis could be used to identify the tissue origin of plasma DNA derived from the tumor. Circulating DNA analysis has increasingly been used in the management of cancer patients, in particular for guiding the use of target therapy and monitoring disease progression (*Mok et al., 2017*; *Wan et al., 2020*; *Yung et al., 2009*). Recently, it has been demonstrated that the analysis for cancer-derived DNA in plasma is useful for the screening of cancers in asymptomatic individuals (*Chan et al., 2017*; *Lennon et al., 2020*). As genetic and methylation aberrations are present in almost all cancers, the detection of cancer-associated alterations in plasma DNA can potentially serve as a universal tumor marker for a wide variety of cancers. The capability of a tumor marker for picking up multiple types of cancers can greatly enhance the cost-effectiveness of a cancer screening program. However, the lack of tissue or organ specificity of these tests also poses practical challenges on the workup of subjects with positive test results. In the screening study by Lennon et al., subjects tested positive were further investigated with PET-CT to confirm and localize a possible tumor (*Lennon et al., 2020*). However, if the tissue origin and location can be obtained from the ctDNA analysis, more targeted investigation on the potentially affected organ may be performed. For example, a colonoscopy can be performed for individuals who are suspected of having colorectal cancers. This targeted investigation approach not only provides a more accurate assessment for cancers, it also reduces the radiation exposure of the tested positive subjects. Here, we used the GETMap analysis to determine the tissue origin of plasma DNA carrying cancer-associated mutations. First, we compared the sequencing results of the tumor tissues and the blood cells to identify the mutations in the tumor tissues of two HCC patients. In contrast to the pregnancy and transplantation models which used microarray for genotyping, we used whole-genome sequencing to identify the cancer-associated mutations as these mutations would not be covered by the whole-genome arrays. In our simulation analysis, the numbers of informative SNPs and mutations identified are shown to provide a median accuracy of 98.3%. After bisulfite sequencing of the plasma DNA, DNA molecules carrying the cancer-associated mutations were identified and their methylation profiles were used to deduce the contribution from different tissues. The liver was deduced to be the key contributor to these cancer-derived plasma DNA molecules with a contribution of 90% and 87% for the two male HCC patients. The remaining portion, that is, 10% and 13%, were attributed to placental contribution. The attribution of a small proportion of ctDNA to originate from the placenta may be due to the fact that global hypomethylation and hypermethylation of tumor suppressor genes are common features in both the placenta and tumor tissues (*Chan et al., 2013a*;

Feinberg and Vogelstein, 1983; *Lun et al., 2013*). Although this analysis suggests that GETMap analysis may be useful for revealing the tissue origin of ctDNA, the requirement of tumor tissues for mutation identification limits its practical application in cancer screening. To overcome this, we further attempted to identify cancer mutations directly from plasma DNA sequencing. In this regard, non-bisulfite sequencing for the plasma DNA and the blood cells of the cancer patients were performed. The single nucleotide variants present in the plasma DNA but not in the blood cells were regarded as cancer-associated mutations. GETMap analysis was performed on the plasma DNA molecules carrying these mutations using the bisulfite sequencing data. Despite a smaller number of cancer-associated mutations could be identified by directly sequencing plasma DNA compared with sequencing the tumor tissues, the liver was again correctly identified as the key contributor to these cancer-derived DNA molecules. These results suggest that the GETMap analysis could be useful in revealing the tissue origin and location of a concealed cancer in patients who are screened positive with a tumor marker that detects various types of cancers.

We further challenged GETMap analysis with a complex scenario where a woman developed lymphoma during pregnancy. Her plasma consisted of DNA derived from the lymphoma tissues, the fetus, and the normal cells. As fetal tissue was not available, fetal genotypes were deduced by sequencing plasma DNA, maternal blood cells/buccal cells, and tumor tissues. Sequence variants present in plasma that overlap with the dbSNP database but absent in the tumor tissues were regarded as fetal-specific alleles. Variants detected in plasma and tumor tissues, but not overlapping with the dbSNP database were regarded as tumor-specific. Plasma DNA molecules carrying these fetal-specific alleles were deduced to be predominantly (95%) derived from the placenta, whereas those carrying the tumor-specific alleles were solely from lymphocytes.

There has been increasing interest in the tissue composition circulating cell-free DNA. Methods based on analysis of DNA methylation (*Gai et al., 2018*; *Lehmann-Werman et al., 2016*; *Sun et al., 2015*), nucleosome footprint (*Snyder et al., 2016*; *Sun et al., 2019*), sequence motifs, end coordinates, and jaggedness (*Chan et al., 2016*; *Jiang et al., 2018*; *Jiang et al., 2020a*; *Jiang et al., 2020b*) have been developed. However, existing methods only allow the deconvolution of all the DNA as a single entity. In contrast, GETMap analysis can determine the tissue origin of subsets of plasma DNA that carry different genetic variations. The specific analysis of a particular component can enhance the signal-to-noise ratio and eliminate the variation caused by the difference in the concentrations of the target DNA, for example, DNA derived from the tumor. Furthermore, clonal hematopoiesis has been identified as one important source of false-positive results for liquid biopsy-based cancer screening tests. In this regard, GETMap would be useful for identifying the hematopoietic origin of the abnormal signal in such cases. Although the number of cases is relatively small in this proof-of-principle study, we have illustrated the potential applications in cancer detection, prenatal testing, and organ transplant monitoring. As the current format of this method is based on whole-genome bisulfite sequencing, identification of cytosine to thymine alteration is less efficient because bisulfite treatment would convert unmethylated cytosine to thymine. A targeted sequencing approach enriching for regions with mutation hotspots and differential methylation across different tissues can be developed to enhance the cost-effectiveness of this approach.

Materials and methods

Samples and processing

The project was approved by the Joint Chinese University of Hong Kong-Hospital Authority New Territories East Cluster Clinical Research Ethics Committee (approval reference number 2011.204). All participants provided written informed consent. Pregnant women and HCC patients were recruited from the Prince of Wales Hospital of Hong Kong. The pregnant woman with lymphoma was recruited from the Hong Kong Sanatorium and Hospital, Hong Kong. Lung transplant recipients were recruited from the National Institutes of Health (NIH) (iRIS reference number 363880). Plasma samples were collected longitudinally at one or several time points after transplantation. Venous blood samples were collected into EDTA-containing tubes and centrifuged at 1600 *g* for 10 min. The plasma portion was recentrifuged at 16,000 *g* to remove residual blood cells. DNA from plasma was extracted with the QIAamp Circulating Nucleic Acid Kit (Qiagen).

Identification of tumor-specific mutations in HCC patients

We prepared libraries using DNA extracted from the tumor tissue and buffy coat with the TruSeq Nano DNA Library Prep Kit (Illumina). Paired-end (2 × 75 bp) sequencing was performed on the HiSeq4000 system (Illumina). Sequencing data were aligned to the human reference genome using the Burrows-Wheeler Aligner (*Li and Durbin, 2010*). We compared the data of tumor tissue with that of buffy coat to call the tumor-specific mutations using the Genome Analysis Toolkit (version 4.1.2.0) (*McKenna et al., 2010*).

To call the tumor-specific mutations directly from the plasma, DNA isolated from the plasma was submitted to library preparation and sequencing. The sequencing data of plasma DNA were then compared with that of the buffy coat to identify the tumor-specific mutations. Single nucleotides variations observed in plasma for more than a threshold number of occasions but not in the buffy coat were identified as candidate mutations. The threshold was based on the total number of sequenced reads covering the variant's nucleotide position as described in our previous study (*Chan et al., 2016*). In addition, the sequencing reads covering these candidate mutations were realigned to the reference human genome using a second alignment software which could reduce the number of false-positive results caused by alignment errors as described previously (*Chan et al., 2016*).

Identification of tumor-specific mutations and fetal-specific SNPs in the pregnant women with lymphoma

The DNA extracted from the maternal plasma, tumor cells, and normal cells were submitted to library preparation using either the KAPA HTP Library Preparation Kit (Kapa Biosystems) or the TruSeq Nano DNA Library Prep Kit (Illumina) following the manufacturer's instructions. The 2 × 75 (paired-end mode) cycles of sequencing were performed using the Illumina platforms, including the HiSeq and NextSeq. To call the plasma-specific variants, we compared the sequencing data of DNA extracted from the maternal plasma with that from the normal cells using the dynamic cutoff algorithm as described previously (*Chan et al., 2016*). We used the biallelic SNPs downloaded from the dbSNP database (Build 135) to classify the plasma-specific variants. For plasma-specific variants within the dbSNP database, we further filtered out the variants that present in the tumor tissue to obtain the fetal-specific SNPs. For the non-dbSNP variants, the single nucleotide variants observed in at least three molecules from the tumor tissue sequencing data were remained as tumor-specific variants. The bioinformatic pipeline for filtering these mutations was written in Python script.

Microarray-based genotyping

Pre-transplant blood samples were collected from the donor and recipient. Genomic DNA was extracted from whole blood with the DNeasy Blood and Tissue Kit (Qiagen) and amplified with REPLI-g Mini Kit (Qiagen). For the pregnant case, genomic DNA of the mother and fetus were extracted from maternal buffy coat and fetal placenta tissue with the QIAamp DNA Mini Kit (Qiagen). Genotyping was performed on Illumina whole-genome arrays (HumanOmni2.5 or HumanOmni1) following the manufacturer's protocol (*De Vlaminck et al., 2014*).

Bisulfite-treated DNA libraries preparation and sequencing analysis

Libraries were prepared from plasma DNA with the TruSeq Nano DNA Library Prep Kit (Illumina). DNA libraries were subjected to two rounds of bisulfite modification with the EpiTect Bisulfite Kit (Qiagen) following by 12 cycles of PCR amplification. Bisulfite-treated libraries were sequenced in paired-end mode (2 × 75 bp) on a HiSeq 4000 system (Illumina). The sequencing reads were trimmed to remove adapter sequences and low-quality bases (i.e., quality score <5). The trimmed reads were aligned to the human reference genome build hg19 with Methy-Pipe (*Jiang et al., 2014*).

GETMap analysis

The reference methylomes included the whole-genome bisulfite sequencing data of five different tissues, including neutrophils, lymphocytes (combining B and T lymphocytes), liver, and lung from the BLUEPRINT Project (*Martens and Stunnenberg, 2013*), Roadmap Epigenomics (*Roadmap Epigenomics Consortium et al., 2015*), ENCODE (*Davis et al., 2018*), and GEO (*Barrett et al., 2013*). In addition, bisulfite sequencing data of two placenta tissues generated by

our group were used as tissue-specific methylomes. The sequencing reads were aligned to the human reference genome build hg19 with bwa-meth (https://github.com/brentp/bwa-meth). After alignment, the methylation levels for 28,217,006 CpG sites across five types of tissues were determined. CpG sites fulfilling the following criteria were used for the analysis: (i) in the five reference tissues, the difference between the highest and lowest methylation levels was greater than 25% and (ii) after removing either tissue with the highest or the lowest methylation level, the coefficient of variation of methylation level across the remaining reference tissues was less than 0.3. We retrieved the methylation levels of different tissues across the set of CpG sites covered by the set of DNA molecules carrying the genetic variants. The measured CpG methylation levels of DNA molecules were recorded in a vector (X) and the retrieved reference methylation levels across different tissues were recorded in a matrix (M). The proportional contributions (P) from different tissues to donor- or recipient-specific DNA molecules were deduced by quadratic programming:

$$\bar{X}_i = \sum_k (p_k \times M_{ik}),$$

where $\bar{X}_i$ represents the methylation density of a CpG site i in the DNA mixture; p_k represents the proportional contribution of cell type k to the DNA mixture; M_{ik} represents the methylation density of the CpG site i in the cell type k. When the number of sites is the same or larger than the number of organs, the values of individual p_k could be determined.

The aggregated contribution of all cell types would be constrained to be 100%:

$$\sum_k p_k = 100\%$$

Furthermore, all the organs' contributions would be required to be non-negative:

$$p_k \geq 0, \forall k$$

The GETMap deconvolution analysis was performed with a program written in Python (http://www.python.org/).

Sample information

The information of all the samples analyzed in this study, including sequencing depth, number of informative SNPs, number of informative sequencing fragments, number of informative CpG sites, and number of CpG sites used for deconvolution, are provided in *Supplementary file 1*.

Additional information

Competing interests

YM Dennis Lo: Reviewing editor, eLife. Holds equities in DRA, Take2 and Grail. Serves as a scientific cofounder and consultant of Grail. Receives research funding from Grail. Receives royalties from Grail, Illumina, Sequenom, DRA, Take2 and Xcelom. Filed a patent application (US15/214,998). Peiyong Jiang: Holds equities in Grail. Serves as a director of KingMed Future. Received patent royalties from Grail, Illumina, Sequenom, DRA, Take2 and Xcelom. Filed a patent application (US15/214,998). Rossa WK Chiu: Holds equities in DRA, Take2 and Grail. Is a consultant to Grail and Illumina. Receives research funding from Grail. Receives royalties from Grail, Illumina, Sequenom, DRA, Take2 and Xcelom. Filed a patent application (US15/214,998). KC Allen Chan: Holds equities in DRA, Take2 and Grail. Is a consultant to and receives research funding from Grail. Receives royalties

Author ORCIDs

Xiaodan Fan (iD) http://orcid.org/0000-0002-2744-9030
KC Allen Chan (iD) https://orcid.org/0000-0003-1780-1691
YM Dennis Lo (iD) https://orcid.org/0000-0001-8746-0293

Ethics

Human subjects: The project was approved by the Joint Chinese University of Hong Kong-Hospital Authority New Territories East Cluster Clinical Research Ethics Committee (approval reference number 2011.204). All participants provided written informed consent.

Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.64356.sa1
Author response https://doi.org/10.7554/eLife.64356.sa2

Additional files

Supplementary files
• Supplementary file 1. The information of all the samples analyzed in this study, including sequencing depth, number of informative single nucleotide polymorphisms (SNPs), number of informative sequencing fragments, number of informative CpG sites, and number of CpG sites used for deconvolution.

• Transparent reporting form

Data availability
Sequencing data have been deposited in EGA under the accession code EGAS00001004788.

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Gai W, Zhou Z, Jiang P, Cheng SH, Chiu RWK, Chan KCA, Lo YMD	2021	Methylation analysis for plasma DNA of patients with organ transplantation	https://www.ebi.ac.uk/ega/studies/EGAS00001004788	The European Genome-phenome Archive, EGAS00001004788

References

Alberry M, Maddocks D, Jones M, Abdel Hadi M, Abdel-Fattah S, Avent N, Soothill PW. 2007. Free fetal DNA in maternal plasma in anembryonic pregnancies: confirmation that the origin is the trophoblast. *Prenatal Diagnosis* 27:415–418. DOI: https://doi.org/10.1002/pd.1700, PMID: 17286310

Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A. 2013. NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Research* 41:D991–D995. DOI: https://doi.org/10.1093/nar/gks1193, PMID: 23193258

CCGA Consortium, Liu MC, Oxnard GR, Klein EA, Swanton C, Seiden MV. 2020. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Annals of Oncology* 31:745–759. DOI: https://doi.org/10.1016/j.annonc.2020.02.011, PMID: 33506766

Chan KC, Jiang P, Chan CW, Sun K, Wong J, Hui EP, Chan SL, Chan WC, Hui DS, Ng SS, Chan HL, Wong CS, Ma BB, Chan AT, Lai PB, Sun H, Chiu RW, Lo YM. 2013a. Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing. *PNAS* 110:18761–18768. DOI: https://doi.org/10.1073/pnas.1313995110, PMID: 24191000

Chan KC, Jiang P, Zheng YW, Liao GJ, Sun H, Wong J, Siu SS, Chan WC, Chan SL, Chan AT, Lai PB, Chiu RW, Lo YM. 2013b. Cancer genome scanning in plasma: detection of tumor-associated copy number aberrations, single-nucleotide variants, and tumoral heterogeneity by massively parallel sequencing. *Clinical Chemistry* 59:211–224. DOI: https://doi.org/10.1373/clinchem.2012.196014, PMID: 23065472

Chan KC, Jiang P, Sun K, Cheng YK, Tong YK, Cheng SH, Wong AI, Hudecova I, Leung TY, Chiu RW, Lo YM. 2016. Second generation noninvasive fetal genome analysis reveals de novo mutations, single-base parental inheritance, and preferred DNA ends. *PNAS* 113:E8159–E8168. DOI: https://doi.org/10.1073/pnas.1615800113, PMID: 27799561

Chan KCA, Woo JKS, King A, Zee BCY, Lam WKJ, Chan SL, Chu SWI, Mak C, Tse IOL, Leung SYM, Chan G, Hui EP, Ma BBY, Chiu RWK, Leung S-F, van Hasselt AC, Chan ATC, Lo YMD. 2017. Analysis of plasma Epstein–Barr Virus DNA to Screen for Nasopharyngeal Cancer. *New England Journal of Medicine* 377:513–522. DOI: https://doi.org/10.1056/NEJMoa1701717

Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, Hilton JA, Jain K, Baymuradov UK, Narayanan AK, Onate KC, Graham K, Miyasato SR, Dreszer TR, Strattan JS, Jolanki O, Tanaka FY, Cherry JM. 2018. The encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Research* 46:D794–D801. DOI: https://doi.org/10.1093/nar/gkx1081, PMID: 29126249

De Vlaminck I, Valantine HA, Snyder TM, Strehl C, Cohen G, Luikart H, Neff NF, Okamoto J, Bernstein D, Weisshaar D, Quake SR, Khush KK. 2014. Circulating cell-free DNA enables noninvasive diagnosis of heart transplant rejection. *Science Translational Medicine* 6:241ra77. DOI: https://doi.org/10.1126/scitranslmed.3007803, PMID: 24944192

De Vlaminck I, Martin L, Kertesz M, Patel K, Kowarsky M, Strehl C, Cohen G, Luikart H, Neff NF, Okamoto J, Nicolls MR, Cornfield D, Weill D, Valantine H, Khush KK, Quake SR. 2015. Noninvasive monitoring of infection and rejection after lung transplantation. *PNAS* 112:13336–13341. DOI: https://doi.org/10.1073/pnas.1517494112, PMID: 26460048

Feinberg AP, Vogelstein B. 1983. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 301:89–92. DOI: https://doi.org/10.1038/301089a0, PMID: 6185846

Gai W, Ji L, Lam WKJ, Sun K, Jiang P, Chan AWH, Wong J, Lai PBS, Ng SSM, Ma BBY, Wong GLH, Wong VWS, Chan HLY, Chiu RWK, Lo YMD, Chan KCA. 2018. Liver- and Colon-Specific DNA methylation markers in plasma for investigation of colorectal cancers with or without liver metastases. *Clinical Chemistry* **64**:1239–1249. DOI: https://doi.org/10.1373/clinchem.2018.290304, PMID: 29903871

Gasteiger G, Fan X, Dikiy S, Lee SY, Rudensky AY. 2015. Tissue residency of innate lymphoid cells in lymphoid and nonlymphoid organs. *Science* **350**:981–985. DOI: https://doi.org/10.1126/science.aac9593, PMID: 26472762

Jiang P, Sun K, Lun FM, Guo AM, Wang H, Chan KC, Chiu RW, Lo YM, Sun H. 2014. Methy-Pipe: an integrated bioinformatics pipeline for whole genome bisulfite sequencing data analysis. *PLOS ONE* **9**:e100360. DOI: https://doi.org/10.1371/journal.pone.0100360, PMID: 24945300

Jiang P, Sun K, Tong YK, Cheng SH, Cheng THT, Heung MMS, Wong J, Wong VWS, Chan HLY, Chan KCA, Lo YMD, Chiu RWK. 2018. Preferred end coordinates and somatic variants as signatures of circulating tumor DNA associated with hepatocellular carcinoma. *PNAS* **115**:E10925–E10933. DOI: https://doi.org/10.1073/pnas.1814616115, PMID: 30373822

Jiang P, Sun K, Peng W, Cheng SH, Ni M, Yeung PC, Heung MMS, Xie T, Shang H, Zhou Z, Chan RWY, Wong J, Wong VWS, Poon LC, Leung TY, Lam WKJ, Chan JYK, Chan HLY, Chan KCA, Chiu RWK, et al. 2020a. Plasma DNA End-Motif profiling as a fragmentomic marker in Cancer, pregnancy, and transplantation. *Cancer Discovery* **10**:664–673. DOI: https://doi.org/10.1158/2159-8290.CD-19-0622, PMID: 32111602

Jiang P, Xie T, Ding SC, Zhou Z, Cheng SH, Chan RWY, Lee W-S, Peng W, Wong J, Wong VWS, Chan HLY, Chan SL, Poon LCY, Leung TY, Chan KCA, Chiu RWK, Lo YMD. 2020b. Detection and characterization of jagged ends of double-stranded DNA in plasma. *Genome Research* **30**::1144–1153. DOI: https://doi.org/10.1101/gr.261396.120

Kitzman JO, Snyder MW, Ventura M, Lewis AP, Qiu R, Simmons LE, Gammill HS, Rubens CE, Santillan DA, Murray JC, Tabor HK, Bamshad MJ, Eichler EE, Shendure J. 2012. Noninvasive Whole-Genome sequencing of a human fetus. *Science Translational Medicine* **4**:137ra76. DOI: https://doi.org/10.1126/scitranslmed.3004323

Knight SR, Thorne A, Lo Faro ML. 2019. Donor-specific Cell-free DNA as a biomarker in solid organ transplantation A systematic review. *Transplantation* **103**:273–283. DOI: https://doi.org/10.1097/TP.0000000000002482, PMID: 30308576

Koh W, Pan W, Gawad C, Fan HC, Kerchner GA, Wyss-Coray T, Blumenfeld YJ, El-Sayed YY, Quake SR. 2014. Noninvasive in vivo monitoring of tissue-specific global gene expression in humans. *PNAS* **111**:7361–7366. DOI: https://doi.org/10.1073/pnas.1405528111, PMID: 24799715

Leary RJ, Sausen M, Kinde I, Papadopoulos N, Carpten JD, Craig D, O'Shaughnessy J, Kinzler KW, Parmigiani G, Vogelstein B, Diaz LA, Velculescu VE. 2012. Detection of chromosomal alterations in the circulation of Cancer patients with whole-genome sequencing. *Science Translational Medicine* **4**:162ra154. DOI: https://doi.org/10.1126/scitranslmed.3004742, PMID: 23197571

Lehmann-Werman R, Neiman D, Zemmour H, Moss J, Magenheim J, Vaknin-Dembinsky A, Rubertsson S, Nellgård B, Blennow K, Zetterberg H, Spalding K, Haller MJ, Wasserfall CH, Schatz DA, Greenbaum CJ, Dorrell C, Grompe M, Zick A, Hubert A, Maoz M, et al. 2016. Identification of tissue-specific cell death using methylation patterns of circulating DNA. *PNAS* **113**:E1826–E1834. DOI: https://doi.org/10.1073/pnas.1519286113, PMID: 26976580

Lennon AM, Buchanan AH, Kinde I, Warren A, Honushefsky A, Cohain AT, Ledbetter DH, Sanfilippo F, Sheridan K, Rosica D, Adonizio CS, Hwang HJ, Lahouel K, Cohen JD, Douville C, Patel AA, Hagmann LN, Rolston DD, Malani N, Zhou S, et al. 2020. Feasibility of blood testing combined with PET-CT to screen for Cancer and guide intervention. *Science* **369**:eabb9601. DOI: https://doi.org/10.1126/science.abb9601, PMID: 32345712

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**:589–595. DOI: https://doi.org/10.1093/bioinformatics/btp698, PMID: 20080505

Lo YMD, Tein MSC, Pang CCP, Yeung CK, Tong K-L, Hjelm NM, Magnus Hjelm N. 1998. Presence of donor-specific DNA in plasma of kidney and liver-transplant recipients. *The Lancet* **351**:1329–1330. DOI: https://doi.org/10.1016/s0140-6736(05)79055-3

Lo YM, Chan KC, Sun H, Chen EZ, Jiang P, Lun FM, Zheng YW, Leung TY, Lau TK, Cantor CR, Chiu RW. 2010. Maternal plasma DNA sequencing reveals the Genome-Wide genetic and mutational profile of the fetus. *Science Translational Medicine* **2**:61ra91. DOI: https://doi.org/10.1126/scitranslmed.3001720, PMID: 21148127

Lun FM, Chiu RW, Sun K, Leung TY, Jiang P, Chan KC, Sun H, Lo YM. 2013. Noninvasive prenatal methylomic analysis by genomewide bisulfite sequencing of maternal plasma DNA. *Clinical Chemistry* **59**:1583–1594. DOI: https://doi.org/10.1373/clinchem.2013.212274, PMID: 23857673

Martens JH, Stunnenberg HG. 2013. BLUEPRINT: mapping human blood cell epigenomes. *Haematologica* **98**:1487–1489. DOI: https://doi.org/10.3324/haematol.2013.094243, PMID: 24091925

Masuzaki H, Miura K, Yoshiura KI, Yoshimura S, Niikawa N, Ishimaru T. 2004. Detection of cell free placental DNA in maternal plasma: direct evidence from three cases of confined placental mosaicism. *Journal of Medical Genetics* **41**:289–292. DOI: https://doi.org/10.1136/jmg.2003.015784, PMID: 15060106

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**:1297–1303. DOI: https://doi.org/10.1101/gr.107524.110, PMID: 20644199

Mok TS, Wu Y-L, Ahn M-J, Garassino MC, Kim HR, Ramalingam SS, Shepherd FA, He Y, Akamatsu H, Theelen W, Lee CK, Sebastian M, Templeton A, Mann H, Marotti M, Ghiorghiu S, Papadimitrakopoulou VA. 2017.

Osimertinib or Platinum–Pemetrexed in *EGFR* T790M–Positive Lung Cancer. *New England Journal of Medicine* **376**:629–640. DOI: https://doi.org/10.1056/NEJMoa1612674

Moss J, Magenheim J, Neiman D, Zemmour H, Loyfer N, Korach A, Samet Y, Maoz M, Druid H, Arner P, Fu K-Y, Kiss E, Spalding KL, Landesberg G, Zick A, Grinshpun A, Shapiro AMJ, Grompe M, Wittenberg AD, Glaser B, et al. 2018. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. *Nature Communications* **9**:1–12. DOI: https://doi.org/10.1038/s41467-018-07466-6

Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, Wu YC, Pfenning AR, et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* **518**:317–330. DOI: https://doi.org/10.1038/nature14248, PMID: 25693563

Schütz E, Fischer A, Beck J, Harden M, Koch M, Wuensch T, Stockmann M, Nashan B, Kollmar O, Matthaei J, Kanzow P, Walson PD, Brockmöller J, Oellerich M. 2017. Graft-derived cell-free DNA, a noninvasive early rejection and graft damage marker in liver transplantation: a prospective, observational, multicenter cohort study. *PLOS Medicine* **14**:e1002286. DOI: https://doi.org/10.1371/journal.pmed.1002286, PMID: 28441386

Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J. 2016. Cell-free DNA comprises an in vivo nucleosome footprint that informs its Tissues-Of-Origin. *Cell* **164**:57–68. DOI: https://doi.org/10.1016/j.cell.2015.11.050, PMID: 26771485

Sun K, Jiang P, Chan KC, Wong J, Cheng YK, Liang RH, Chan WK, Ma ES, Chan SL, Cheng SH, Chan RW, Tong YK, Ng SS, Wong RS, Hui DS, Leung TN, Leung TY, Lai PB, Chiu RW, Lo YM. 2015. Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, Cancer, and transplantation assessments. *PNAS* **112**:E5503–E5512. DOI: https://doi.org/10.1073/pnas.1508736112, PMID: 26392541

Sun K, Jiang P, Cheng SH, Cheng THT, Wong J, Wong VWS, Ng SSM, Ma BBY, Leung TY, Chan SL, Mok TSK, Lai PBS, Chan HLY, Sun H, Chan KCA, Chiu RWK, Lo YMD. 2019. Orientation-aware plasma cell-free DNA fragmentation analysis in open chromatin regions informs tissue of origin. *Genome Research* **29**:418–427. DOI: https://doi.org/10.1101/gr.242719.118, PMID: 30808726

Tsui NB, Jiang P, Wong YF, Leung TY, Chan KC, Chiu RW, Sun H, Lo YM. 2014. Maternal plasma RNA sequencing for genome-wide transcriptomic profiling and identification of pregnancy-associated transcripts. *Clinical Chemistry* **60**:954–962. DOI: https://doi.org/10.1373/clinchem.2014.221648, PMID: 24740362

Wan JCM, Heider K, Gale D, Murphy S, Fisher E, Mouliere F, Ruiz-Valdepenas A, Santonja A, Morris J, Chandrananda D, Marshall A, Gill AB, Chan PY, Barker E, Young G, Cooper WN, Hudecova I, Marass F, Mair R, Brindle KM, et al. 2020. ctDNA monitoring using patient-specific sequencing and integration of variant reads. *Science Translational Medicine* **12**:eaaz8084. DOI: https://doi.org/10.1126/scitranslmed.aaz8084, PMID: 32554709

Wong IH, Lo YM, Zhang J, Liew CT, Ng MH, Wong N, Lai PB, Lau WY, Hjelm NM, Johnson PJ. 1999. Detection of aberrant p16 methylation in the plasma and serum of liver Cancer patients. *Cancer Research* **59**:71–73. PMID: 9892188

Yung TK, Chan KC, Mok TS, Tong J, To KF, Lo YM. 2009. Single-molecule detection of epidermal growth factor receptor mutations in plasma by microfluidics digital PCR in non-small cell lung Cancer patients. *Clinical Cancer Research* **15**:2076–2084. DOI: https://doi.org/10.1158/1078-0432.CCR-08-2622, PMID: 19276259