

**Title:** Genetic-epigenetic tissue mapping for plasma DNA: applications in prenatal testing, transplantation and oncology

**Authors:** Wanxia Gai<sup>1,2</sup>, Ze Zhou<sup>1,2</sup>, Sean Agbor-Enoh<sup>3,4,5</sup>, Xiaodan Fan<sup>6</sup>, Sheng Lian<sup>6</sup>, Peiyong Jiang<sup>1,2</sup>, Suk Hang Cheng<sup>1,2</sup>, John Wong<sup>7</sup>, Stephen L. Chan<sup>8</sup>, Moon Kyoo Jang<sup>3,5</sup>, Yanqin Yang<sup>3,5</sup>, Raymond H. S. Liang<sup>9</sup>, Wai Kong Chan<sup>10</sup>, Edmond S. K. Ma<sup>10</sup>, Tak Y. Leung<sup>11</sup>, Rossa W. K. Chiu<sup>1,2</sup>, Hannah Valantine<sup>3,5</sup>, K. C. Allen Chan<sup>1,2</sup>, Y. M. Dennis Lo<sup>1,2</sup>

**Affiliations:**

<sup>1</sup>Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, China.

<sup>2</sup>Department of Chemical Pathology, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, New Territories, Hong Kong SAR, China.

<sup>3</sup>Genomic Research Alliance for Transplantation (GRAFT), 10 Center Drive, 7S261, Bethesda, MD 20982, United States

<sup>4</sup>Division of Pulmonary and Critical Care Medicine, The Johns Hopkins School of Medicine, 1830 East Monument Street, Baltimore, MD, United States

<sup>5</sup>Division of Intramural Research, National Heart, Lung and Blood Institute, 10 Center Drive, 7S261, Bethesda, MD 20982, United States

<sup>6</sup>Department of Statistics, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, China

<sup>7</sup>Department of Surgery, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, New Territories, Hong Kong SAR, China

<sup>8</sup>Department of Clinical Oncology, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, New Territories, Hong Kong SAR, China

<sup>9</sup>Comprehensive Oncology Centre, Hong Kong Sanatorium & Hospital, Hong Kong SAR, China

<sup>10</sup>Department of Pathology, Hong Kong Sanatorium & Hospital, Hong Kong SAR, China

<sup>11</sup>Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, New Territories, Hong Kong SAR, China

**Note:** W. Gai, and Z. Zhou contributed equally to this work.

**Corresponding Author:** Y.M. Dennis Lo, The Chinese University of Hong Kong, Prince of Wales Hospital, 30–32 Ngan Shing Street, Shatin, New Territories, Hong Kong SAR, China.  
Phone: +852-3763-6001; Fax: +852-2636-5090; E-mail: 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 hours 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,

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, 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, Sun, et al., 2020; Jiang, Xie, et al., 2020), and the profiling of RNA transcripts (Koh et al., 2014; Tsui et al., 2014). In these methods, the features of interest, e.g. 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; Liu et al., 2020). As cancer-associated genetic and epigenetic changes are present in virtually all types of cancers (Chan, Jiang, Chan, et al., 2013; Chan, Jiang, Zheng, et al., 2013; 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, e.g. 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 (Liu 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 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 Fig. 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, e.g. those carrying fetal-specific alleles, or cancer-associated mutations or donor-derived alleles, are analyzed. Second, only CpG sites near informative 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.

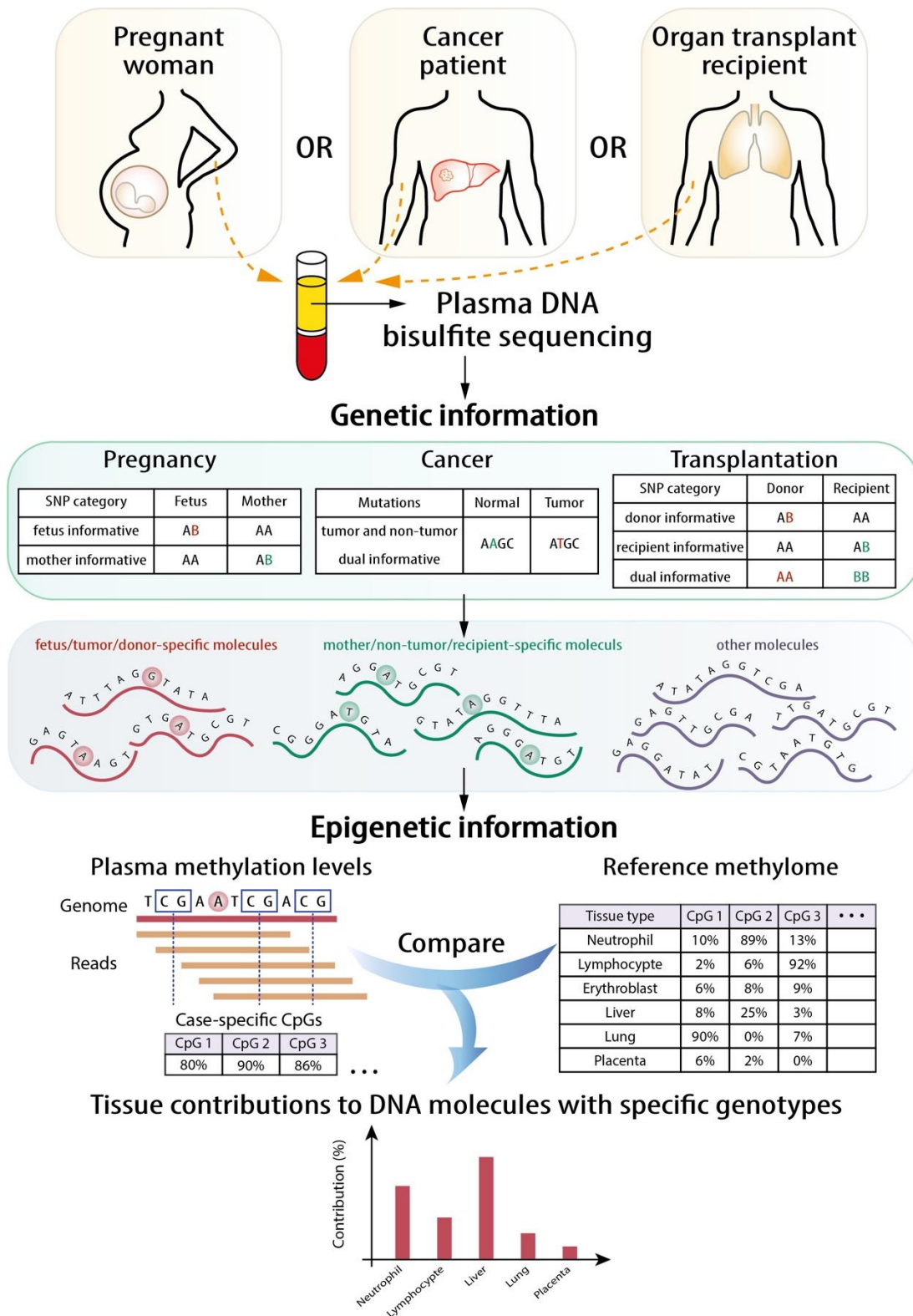


Fig. 1. Schematic illustration of the principle of GETMap analysis. The paired individuals (e.g. fetus/mother and organ donor/recipient, and tumor/normal tissue) are genotyped to identify SNP alleles specific for one of them. After bisulfite sequencing, plasma DNA molecules carrying individual-specific alleles and at least one CpG site are identified. The plasma DNA methylome is 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.



### *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 190,706 (range: 150,168-193,406) maternal-specific informative single nucleotide polymorphisms (SNPs) where the mother was heterozygous and the fetus was homozygous, and a median of 195,331 (range: 146,428-202,800) 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: 8,237-295,671) and 53,039 CpG 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% (Fig. 2A). These results are compatible to the results of previously studies that fetal DNA in maternal plasma is derived from the placenta (Alberry et al., 2007; Masuzaki, 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) (Fig. 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 (Fig. 2C).

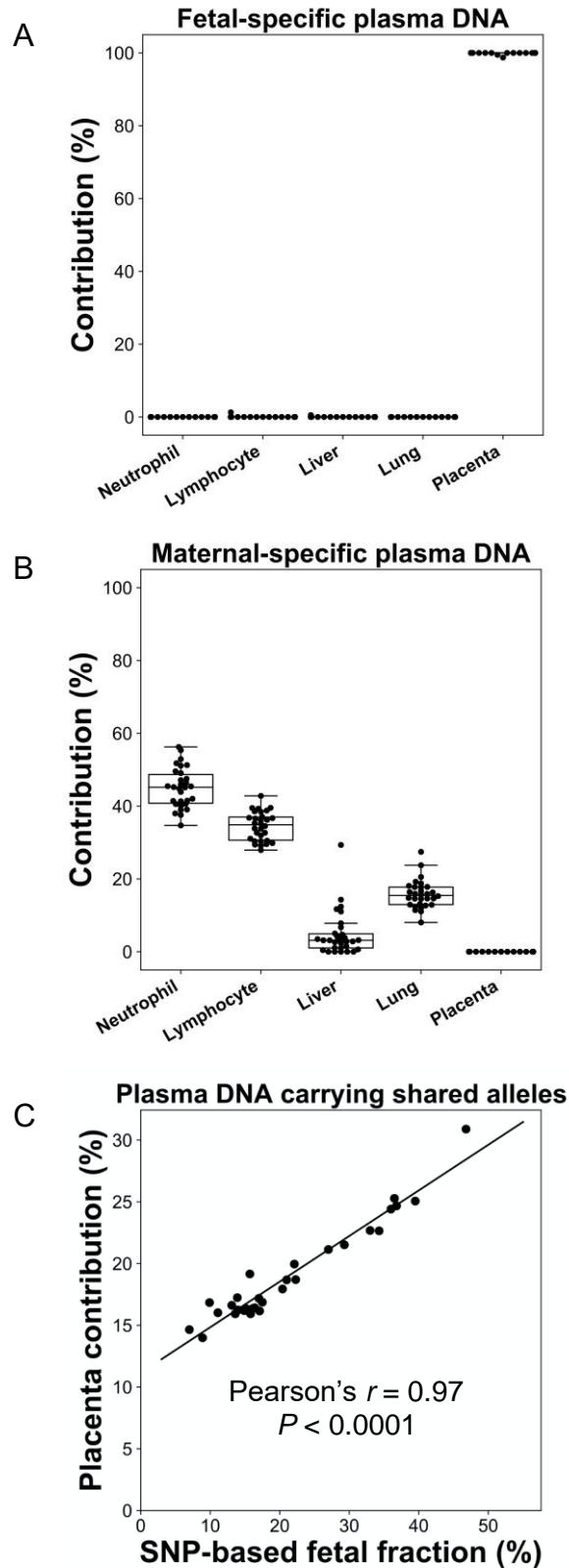


Fig. 2. Percentage contributions of different cell types to maternal plasma DNA carrying (A) fetal-specific alleles; and (B) maternal-specific alleles in 30 pregnant women. (C) Correlation between percentage contribution of the placenta to maternal plasma DNA molecules carrying alleles shared by the fetus and mother and SNP-based fetal DNA fraction.

*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 S1). 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) 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 hours after the transplantation. We performed the GETMap analysis on donor-derived DNA molecules for each sample collected at 72 hours posttransplant (Fig. 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 posttransplant: within 72 hours; in-between 72 hours, 7 days, 10 weeks, and

50 weeks; and beyond 50 weeks. The 40 samples were thus classified into 5 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 (Fig. 3B). The median contributions from the lung to the donor-derived DNA were 17%, 34%, 59%, 51%, 66% for samples in these categories, respectively (Fig. 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, i.e., 78%, 56%, 27%, 41%, and 21% for samples in the 5 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 5 categories, the median contributions of hematopoietic cells were 83%, 86%, 89 %, 94%, and 84%, respectively (Supplementary Fig. S1).

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

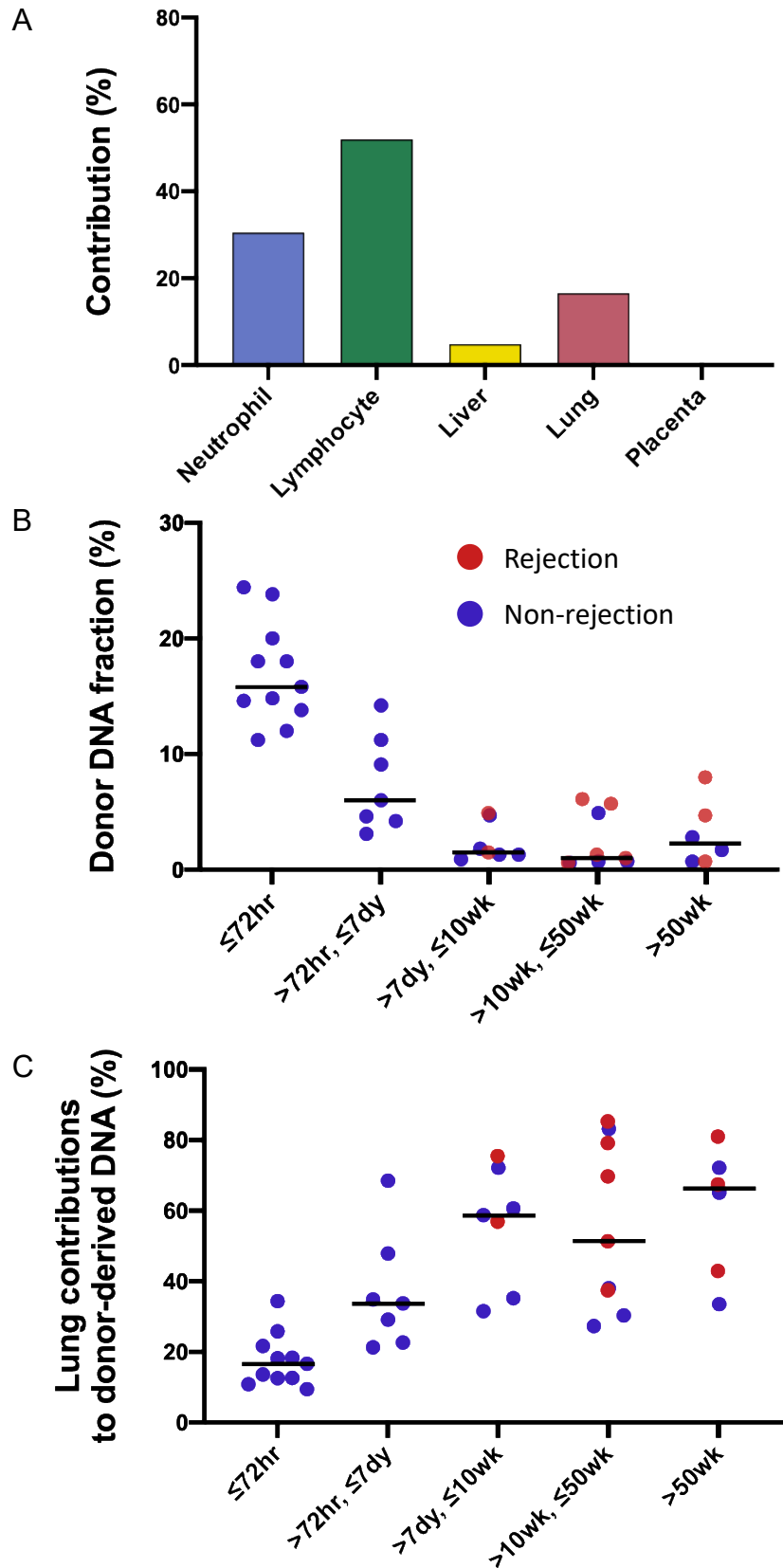


Fig. 3. (A) The median percentage contributions of different cell types to plasma DNA carrying donor-specific alleles in patients with lung transplantation at 72 hours posttransplant. (B) Fractional concentrations of donor-derived DNA and (C) percentage contributions of the lung to plasma DNA carrying donor-specific alleles in patients with lung transplantation.

### *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 hepatocellular cancer (HCC) patients. The two patients were denoted as HCC 1 and 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 6,996 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 number of plasma DNA molecules carrying the mutant alleles were 29,868 and 5,090, and these molecules covered 18,193 and 4,076 CpG sites, respectively. Tissue contributions of these tumor-derived plasma DNA molecules were deduced by GETMap analysis (Fig. 4). 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 number of molecules carrying the wildtype alleles were 153,238 and 26,792, containing 35,883 and 8,156 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).

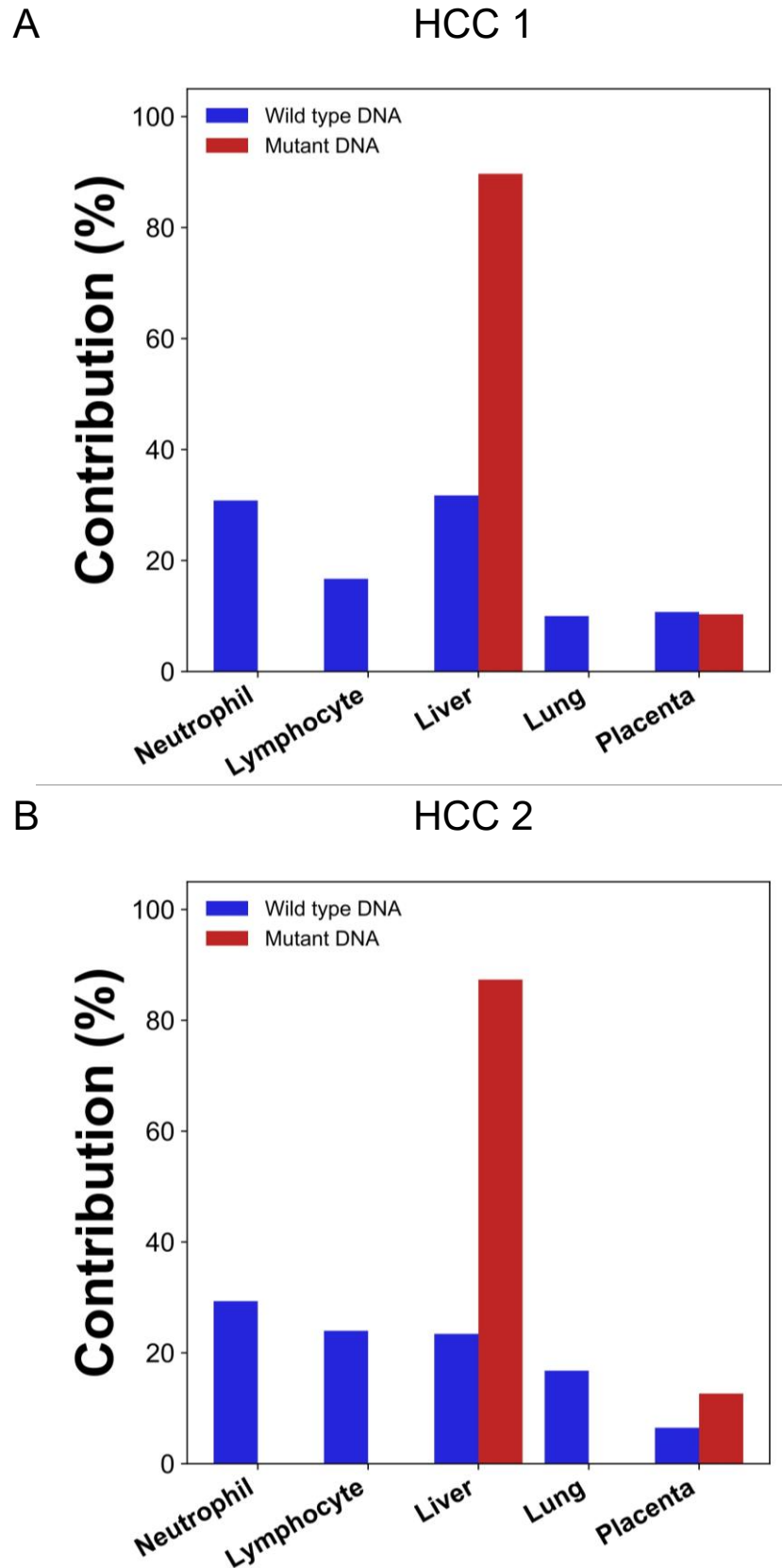


Fig. 4. Percentage contributions of different tissues to plasma DNA with tumor-specific and wildtype alleles in two HCC patients. The tumor-specific mutations were deduced from the tumor tissue.



### *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 50× and 61× haploid genome coverage and those for the buffy coat DNA were 53× and 55× 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 number of candidate mutations identified were 10,864 and 3,446 for the two HCC patients. GETMap analysis was then performed using the plasma DNA bisulfite sequencing data. The number of plasma DNA molecules carrying the cancer mutations were 16,200 and 4,112, and covered 12,887 and 2,991 CpG sites, respectively. For molecules carrying mutations, the contributions from the liver were estimated to be 69% (HCC 1) and 95% (HCC 2) (Fig. 5). 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).

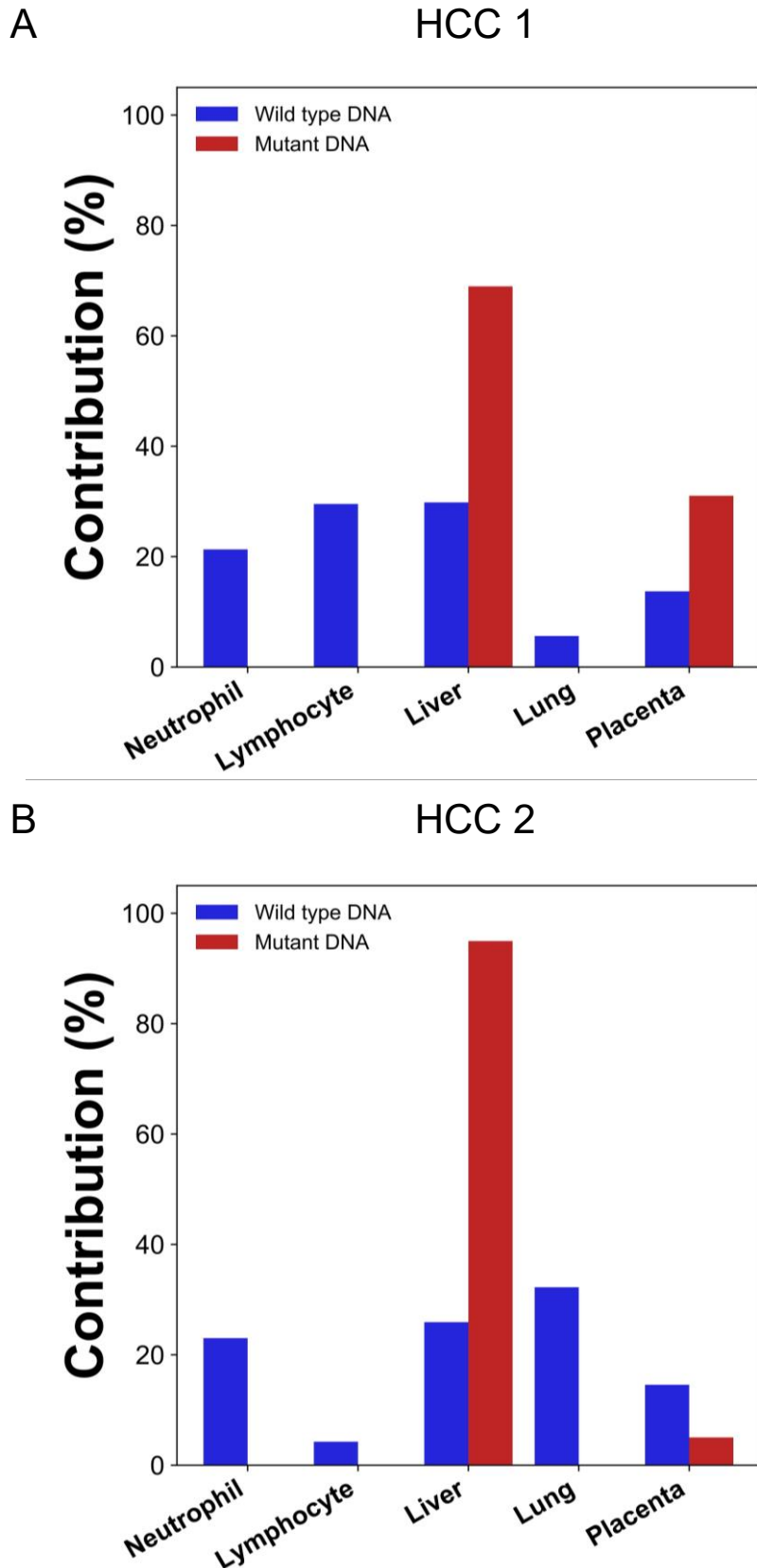


Fig. 5. Percentage contributions of different tissues to plasma DNA with tumor-specific and wildtype alleles in two HCC patients. The tumor-specific mutations were deduced directly from the plasma.

### *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 Supplementary Fig. S2. 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, 2,641 were detected in 3 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 (Supplementary Fig. S3).

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 number of CpG sites covered by the DNA molecules carrying the mutant and wildtype alleles

were 4,781 and 6,660 respectively. For the molecules carrying tumor mutations, it was deduced that 100% was from lymphocytes (Fig. 6A). 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% (Fig. 6B). 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.

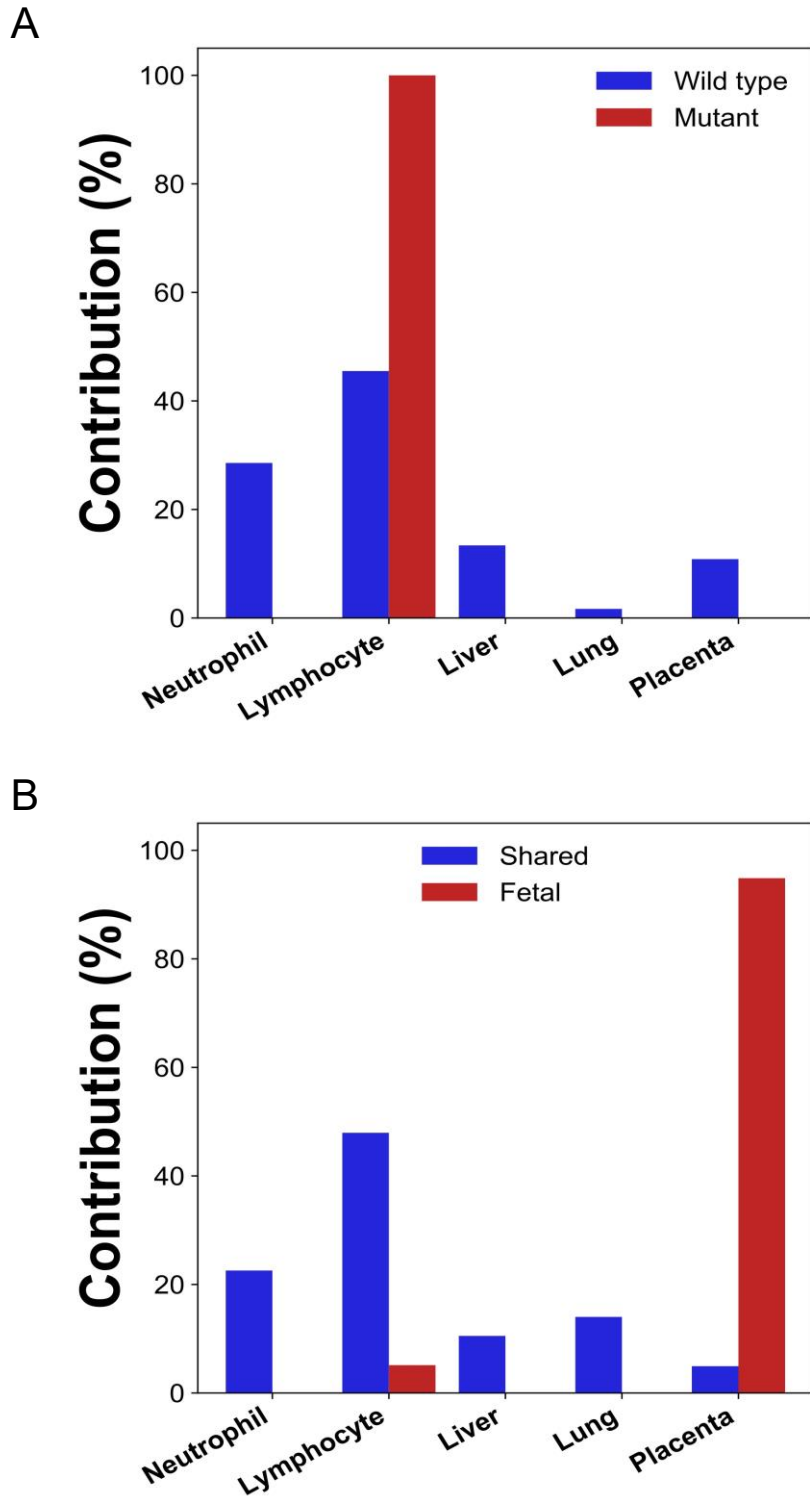


Fig. 6. Percentage contributions of different tissues to (A) plasma DNA with tumor-specific and wildtype alleles, and (B) fetal-specific plasma DNA and DNA carrying the alleles shared by the fetus and the mother in a pregnant woman with lymphoma.

## 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 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 GETMap 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 are 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 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.

Then we 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 Vlamincq 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 hours 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 work up 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. 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 HCC patients. The remaining portion, i.e. 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, Jiang, Chan, et al., 2013; Feinberg & 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 dbSNP database but absent in the tumor tissues were regarded as fetal-specific alleles. Variants detected in plasma and tumor tissues, but not overlap with 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 interests 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, Sun, et al., 2020; Jiang, Xie, et al., 2020) 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, e.g. 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. While the current format of this method is based on whole genome bisulfite sequencing, 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. 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 & Hospital, Hong Kong. Lung transplant recipients were recruited

from the National Institutes of Health (NIH). 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 1,600 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 ( $2 \times 75$  bp) sequencing was performed on the HiSeq4000 system (Illumina). Sequencing data were aligned to the human reference genome using the Burrows-Wheeler Aligner (Li & 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  $2 \times 75$  (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 ( $2 \times 75$  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*

We analyzed whole-genome bisulfite sequencing data of different tissues or cell types from public datasets, including BLUEPRINT Project (Martens & Stunnenberg, 2013), Roadmap Epigenomics (Roadmap Epigenomics Consortium et al., 2015), ENCODE (Davis et al., 2018) and GEO (Barrett et al., 2012) and methylomes generated by our group. The methylation levels for 28,217,006 CpG sites across 6 types of tissues were determined. We retrieved the methylation levels of different tissues across the set of CpG sites covered by donor- and recipient-specific DNA molecules. The measured CpG methylation levels of donor- or recipient-specific 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. To improve the informativeness, the CpG sites showed small variability of methylation levels across all reference tissue types were discarded. The remaining CpG sites were characterized with the coefficient of variation (CV) of methylation levels across different tissues less than 0.3 and a difference between a maximum and a minimum methylation levels among tissues more than 25%.

Additional criteria were included in the algorithm to improve the accuracy. 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/>).

### **Acknowledgement:**

This work was supported by the Research Grants Council of the Hong Kong SAR Government under the Theme-based research scheme (T12-403/15-N and T12-401/16-W), a collaborative research agreement from Grail and the Vice Chancellor's One-Off Discretionary Fund of The Chinese University of Hong Kong (VCF2014021). Y.M.D. Lo is supported by an endowed chair from the Li Ka Shing Foundation.

**Competing interests:**

R.W.K.C., K.C.A.C., and Y.M.D.L. hold equities in DRA, Take2, and Grail. R.W.K.C., K.C.A.C., and Y.M.D.L. are consultants to Grail. R.W.K.C., K.C.A.C., and Y.M.D.L. receive research funding from Grail. Y.M.D.L. is a scientific cofounder of and serves on the scientific advisory board of Grail. R.W.K.C. is a consultant to Illumina. P.J. holds equities in Grail. P.J. is a Director of KingMed Future. P.J., K.C.A.C., R.W.K.C., and Y.M.D.L. received patent royalties from Grail, Illumina, Sequenom, DRA, Take2, and Xcelom. P.J., K.C.A.C., R.W.K.C., and Y.M.D.L. had filed a patent application (US15/214,998).

References:

- Alberry, M., Maddocks, D., Jones, M., Abdel Hadi, M., Abdel-Fattah, S., Avent, N., & Soothill, P. W. (2007). Free fetal DNA in maternal plasma in anembryonic pregnancies: confirmation that the origin is the trophoblast. *Prenatal Diagnosis*, *27*(5), 415–418. <https://doi.org/10.1002/pd.1700>
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C. L., Serova, N., Davis, S., & Soboleva, A. (2012). NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Research*, *41*(D1), D991–D995. <https://doi.org/10.1093/nar/gks1193>
- Chan, K. C. A., Jiang, P., Chan, C. W. M., Sun, K., Wong, J., Hui, E. P., Chan, S. L., Chan, W. C., Hui, D. S. C., Ng, S. S. M., Chan, H. L. Y., Wong, C. S. C., Ma, B. B. Y., Chan, A. T. C., Lai, P. B. S., Sun, H., Chiu, R. W. K., & Lo, Y. M. D. (2013). Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(47), 18761–18768. <https://doi.org/10.1073/pnas.1313995110>
- Chan, K. C. A., Jiang, P., Sun, K., Cheng, Y. K. Y., Tong, Y. K., Cheng, S. H., Wong, A. I. C., Hudecova, I., Leung, T. Y., Chiu, R. W. K., & Lo, Y. M. D. (2016). Second generation noninvasive fetal genome analysis reveals de novo mutations, single-base parental inheritance, and preferred DNA ends. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(50), E8159–E8168. <https://doi.org/10.1073/pnas.1615800113>
- Chan, K. C. A., Jiang, P., Zheng, Y. W. L., Liao, G. J. W. W., Sun, H., Wong, J., Siu, S. S.



- N., Chan, W. C., Chan, S. L., Chan, A. T. C., Lai, P. B. S., Chiu, R. W. K., & Lo, Y. M. D. (2013). 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*(1), 211–224.  
<https://doi.org/10.1373/clinchem.2012.196014>
- Chan, K. C. A., Woo, J. K. S., King, A., Zee, B. C. Y., Lam, W. K. J., Chan, S. L., Chu, S. W. I., Mak, C., Tse, I. O. L., Leung, S. Y. M., Chan, G., Hui, E. P., Ma, B. B. Y., Chiu, R. W. K., Leung, S.-F., van Hasselt, A. C., Chan, A. T. C., & Lo, Y. M. D. (2017). Analysis of plasma Epstein–Barr virus DNA to screen for nasopharyngeal cancer. *New England Journal of Medicine*, *377*(6), 513–522.  
<https://doi.org/10.1056/NEJMoa1701717>
- Davis, C. A., Hitz, B. C., Sloan, C. A., Chan, E. T., Davidson, J. M., Gabdank, I., Hilton, J. A., Jain, K., Baymuradov, U. K., Narayanan, A. K., Onate, K. C., Graham, K., Miyasato, S. R., Dreszer, T. R., Strattan, J. S., Jolanki, O., Tanaka, F. Y., & Cherry, J. M. (2018). The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Research*, *46*(D1), D794–D801. <https://doi.org/10.1093/nar/gkx1081>
- De Vlaminc, I., Martin, L., Kertesz, M., Patel, K., Kowarsky, M., Strehl, C., Cohen, G., Luikart, H., Neff, N. F., Okamoto, J., Nicolls, M. R., Cornfield, D., Weill, D., Valantine, H., Khush, K. K., & Quake, S. R. (2015). Noninvasive monitoring of infection and rejection after lung transplantation. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(43), 13336–13341.  
<https://doi.org/10.1073/pnas.1517494112>
- De Vlaminc, I., Valantine, H. A., Snyder, T. M., Strehl, C., Cohen, G., Luikart, H., Neff, N. F., Okamoto, J., Bernstein, D., Weisshaar, D., Quake, S. R., & Khush, K. K. (2014). Circulating cell-free DNA enables noninvasive diagnosis of heart transplant rejection.

*Science Translational Medicine*, 6(241), 241ra77.

<https://doi.org/10.1126/scitranslmed.3007803>

Feinberg, A. P., & Vogelstein, B. (1983). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature*, 301(5895), 89–92.

<https://doi.org/10.1038/301089a0>

Gai, W., Ji, L., Lam, W. K. J., Sun, K., Jiang, P., Chan, A. W. H., Wong, J., Lai, P. B. S., Ng, S. S. M., Ma, B. B. Y., Wong, G. L. H., Wong, V. W. S., Chan, H. L. Y., Chiu, R. W. K., Lo, Y. M. D., & Chan, K. C. A. (2018). Liver- and colon-specific DNA methylation markers in plasma for investigation of colorectal cancers with or without liver metastases. *Clinical Chemistry*, 64(8), 1239–1249.

<https://doi.org/10.1373/clinchem.2018.290304>

Gasteiger, G., Fan, X., Dikiy, S., Lee, S. Y., & Rudensky, A. Y. (2015). Tissue residency of innate lymphoid cells in lymphoid and nonlymphoid organs. *Science*, 350(6263), 981–985. <https://doi.org/10.1126/science.aac9593>

Jiang, P., Sun, K., Lun, F. M. F., Guo, A. M., Wang, H., Chan, K. C. A., Chiu, R. W. K., Lo, Y. M. D., & Sun, H. (2014). Methy-Pipe: an integrated bioinformatics pipeline for whole genome bisulfite sequencing data analysis. *PLoS ONE*, 9(6), e100360.

<https://doi.org/10.1371/journal.pone.0100360>

Jiang, P., Sun, K., Peng, W., Cheng, S. H., Ni, M., Yeung, P. C., Heung, M. M. S., Xie, T., Shang, H., Zhou, Z., Chan, R. W. Y., Wong, J., Wong, V. W. S., Poon, L. C., Leung, T. Y., Lam, W. K. J., Chan, J. Y. K., Chan, H. L. Y., Chan, K. C. A., ... Lo, Y. M. D. (2020). Plasma DNA end-motif profiling as a fragmentomic marker in cancer, pregnancy, and transplantation. *Cancer Discovery*, 10(5), 664–673.

<https://doi.org/10.1158/2159-8290.CD-19-0622>

Jiang, P., Sun, K., Tong, Y. K., Cheng, S. H., Cheng, T. H. T., Heung, M. M. S., Wong, J.,

- Wong, V. W. S., Chan, H. L. Y., Chan, K. C. A., Lo, Y. M. D., & Chiu, R. W. K. (2018). Preferred end coordinates and somatic variants as signatures of circulating tumor DNA associated with hepatocellular carcinoma. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(46), E10925–E10933. <https://doi.org/10.1073/pnas.1814616115>
- Jiang, P., Xie, T., Ding, S. C., Zhou, Z., Cheng, S. H., Chan, R. W. Y., Lee, W. S., Peng, W., Wong, J., Wong, V. W. S., Chan, H. L. Y., Chan, S. L., Poon, L. C. Y., Leung, T. Y., Chan, K. C. A., Chiu, R. W. K., & Lo, Y. M. D. (2020). Detection and characterization of jagged ends of double-stranded DNA in plasma. *Genome Research*, *30*(8). <https://doi.org/10.1101/gr.261396.120>
- Kitzman, J. O., Snyder, M. W., Ventura, M., Lewis, A. P., Qiu, R., Simmons, L. V. E., Gammill, H. S., Rubens, C. E., Santillan, D. A., Murray, J. C., Tabor, H. K., Bamshad, M. J., Eichler, E. E., & Shendure, J. (2012). Noninvasive whole-genome sequencing of a human fetus. *Science Translational Medicine*, *4*(137), 137ra76. <https://doi.org/10.1126/scitranslmed.3004323>
- Knight, S. R., Thorne, A., & Lo Faro, M. L. (2019). Donor-specific Cell-free DNA as a Biomarker in Solid Organ Transplantation. A Systematic Review. *Transplantation*, *103*(2), 273–283. <https://doi.org/10.1097/TP.0000000000002482>
- Koh, W., Pan, W., Gawad, C., Christina Fan, H., Kerchner, G. A., Wyss-Coray, T., Blumenfeld, Y. J., El-Sayed, Y. Y., & Quake, S. R. (2014). Noninvasive in vivo monitoring of tissue-specific global gene expression in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(20), 7361–7366. <https://doi.org/10.1073/pnas.1405528111>
- Leary, R. J., Sausen, M., Kinde, I., Papadopoulos, N., Carpten, J. D., Craig, D., O’Shaughnessy, J., Kinzler, K. W., Parmigiani, G., Vogelstein, B., Diaz, L. A., &

- Velculescu, V. E. (2012). Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. *Science Translational Medicine*, 4(162), 162ra154. <https://doi.org/10.1126/scitranslmed.3004742>
- 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, M. J., Wasserfall, C. H., Schatz, D. A., Greenbaum, C. J., Dorrell, C., Grompe, M., Zick, A., Hubert, A., ... Dor, Y. (2016). Identification of tissue-specific cell death using methylation patterns of circulating DNA. *Proceedings of the National Academy of Sciences*, 113(13), E1826–E1834. <https://doi.org/10.1073/pnas.1519286113>
- Lennon, A. M., Buchanan, A. H., Kinde, I., Warren, A., Honushefsky, A., Cohain, A. T., Ledbetter, D. H., Sanfilippo, F., Sheridan, K., Rosica, D., Adonizio, C. S., Hwang, H. J., Lahouel, K., Cohen, J. D., Douville, C., Patel, A. A., Hagmann, L. N., Rolston, D. D., Malani, N., ... Papadopoulos, N. (2020). Feasibility of blood testing combined with PET-CT to screen for cancer and guide intervention. *Science (New York, N.Y.)*, 369(6499), eabb9601. <https://doi.org/10.1126/science.abb9601>
- Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, 26(5), 589–595. <https://doi.org/10.1093/bioinformatics/btp698>
- Liu, M. C., Oxnard, G. R., Klein, E. A., Swanton, C., Seiden, M. V., Liu, M. C., Oxnard, G. R., Klein, E. A., Smith, D., Richards, D., Yeatman, T. J., Cohn, A. L., Lapham, R., Clement, J., Parker, A. S., Tummala, M. K., McIntyre, K., Sekeres, M. A., Bryce, A. H., ... Berry, D. A. (2020). Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Annals of Oncology*, 31(6), 745–759. <https://doi.org/10.1016/j.annonc.2020.02.011>
- Lo, Y. M. D., Chan, K. C. A., Sun, H., Chen, E. Z., Jiang, P., Lun, F. M. F., Zheng, Y. W.,

- Leung, T. Y., Lau, T. K., Cantor, C. R., & Chiu, R. W. K. (2010). Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Science Translational Medicine*, 2(61), 61ra91.  
<https://doi.org/10.1126/scitranslmed.3001720>
- Lo, Y. M. D., Tein, M. S. C., Pang, C. C. P., Yeung, C. K., Tong, K. L., & Magnus Hjelm, N. (1998). Presence of donor-specific DNA in plasma of kidney and liver-transplant recipients. *Lancet*, 351(9112), 1329–1330. [https://doi.org/10.1016/S0140-6736\(05\)79055-3](https://doi.org/10.1016/S0140-6736(05)79055-3)
- Lun, F. M. F., Chiu, R. W. K., Sun, K., Leung, T. Y., Jiang, P., Chan, K. C. A., Sun, H., & Lo, Y. M. D. (2013). Noninvasive prenatal methylomic analysis by genomewide bisulfite sequencing of maternal plasma DNA. *Clinical Chemistry*, 59(11), 1583–1594.  
<https://doi.org/10.1373/clinchem.2013.212274>
- Martens, J. H. A., & Stunnenberg, H. G. (2013). BLUEPRINT: mapping human blood cell epigenomes. *Haematologica*, 98(10), 1487–1489.  
<https://doi.org/10.3324/haematol.2013.094243>
- Masuzaki, H. (2004). Detection of cell free placental DNA in maternal plasma: direct evidence from three cases of confined placental mosaicism. *Journal of Medical Genetics*, 41(4), 289–292. <https://doi.org/10.1136/jmg.2003.015784>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303.  
<https://doi.org/10.1101/gr.107524.110>
- Mok, T. S., Wu, Y.-L., Ahn, M.-J., Garassino, M. C., Kim, H. R., Ramalingam, S. S., Shepherd, F. A., He, Y., Akamatsu, H., Theelen, W. S. M. E., Lee, C. K., Sebastian, M.,

- Templeton, A., Mann, H., Marotti, M., Ghiorghiu, S., & Papadimitrakopoulou, V. A. (2017). Osimertinib or Platinum–Pemetrexed in EGFR T790M–Positive Lung Cancer. *New England Journal of Medicine*, *376*(7), 629–640. <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, K. L., Landesberg, G., Zick, A., Grinshpun, A., Shapiro, A. M. J., Grompe, M., Wittenberg, A. D., ... Dor, Y. (2018). Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. *Nature Communications*, *9*(1), 1–12. <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, M. J., Amin, V., Whitaker, J. W., Schultz, M. D., Ward, L. D., Sarkar, A., Quon, G., Sandstrom, R. S., Eaton, M. L., ... Kellis, M. (2015). Integrative analysis of 111 reference human epigenomes. *Nature*, *518*(7539), 317–330. <https://doi.org/10.1038/nature14248>
- Schütz, E., Fischer, A., Beck, J., Harden, M., Koch, M., Wuensch, T., Stockmann, M., Nashan, B., Kollmar, O., Matthaei, J., Kanzow, P., Walson, P. D., 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*(4), e1002286. <https://doi.org/10.1371/journal.pmed.1002286>
- Snyder, M. W., Kircher, M., Hill, A. J., Daza, R. M., & Shendure, J. (2016). Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. *Cell*, *164*(1–2), 57–68. <https://doi.org/10.1016/j.cell.2015.11.050>
- Sun, K., Jiang, P., Chan, K. C. A., Wong, J., Cheng, Y. K. Y., Liang, R. H. S., Chan, W., Ma,

- E. S. K., Chan, S. L., Cheng, S. H., Chan, R. W. Y., Tong, Y. K., Ng, S. S. M., Wong, R. S. M., Hui, D. S. C., Leung, T. Y., Leung, T. N., Lai, P. B. S., Chiu, R. W. K., & Lo, Y. M. D. (2015). Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments. *Proceedings of the National Academy of Sciences*, *112*(40), E5503–E5512.  
<https://doi.org/10.1073/pnas.1508736112>
- Sun, K., Jiang, P., Cheng, S. H., Cheng, T. H. T., Wong, J., Wong, V. W. S., Ng, S. S. M., Ma, B. B. Y., Leung, T. Y., Chan, S. L., Mok, T. S. K., Lai, P. B. S., Chan, H. L. Y., Sun, H., Chan, K. C. A., Chiu, R. W. K., & Lo, Y. M. D. (2019). Orientation-aware plasma cell-free DNA fragmentation analysis in open chromatin regions informs tissue of origin. *Genome Research*, *29*(3), 418–427. <https://doi.org/10.1101/gr.242719.118>
- Tsui, N. B. Y., Jiang, P., Wong, Y. F., Leung, T. Y., Allen Chan, K. C., Chiu, R. W. K., Sun, H., & Dennis Lo, Y. M. (2014). Maternal plasma RNA sequencing for genome-wide transcriptomic profiling and identification of pregnancy-associated transcripts. *Clinical Chemistry*, *60*(7), 954–962. <https://doi.org/10.1373/clinchem.2014.221648>
- Wan, J. C. M., Heider, K., Gale, D., Murphy, S., Fisher, E., Mouliere, F., Ruiz-Valdepenas, A., Santonja, A., Morris, J., Chandrananda, D., Marshall, A., Gill, A. B., Chan, P. Y., Barker, E., Young, G., Cooper, W. N., Hudecova, I., Marass, F., Mair, R., ... Rosenfeld, N. (2020). ctDNA monitoring using patient-specific sequencing and integration of variant reads. *Science Translational Medicine*, *12*(548).  
<https://doi.org/10.1126/scitranslmed.aaz8084>
- Wong, I. H. N., Lo, Y. M. D., Zhang, J., Liew, C. T., Ng, M. H. L., Wong, N., Lai, P. B. S., Lau, W. Y., Hjelm, N. M., & Johnson, P. J. (1999). Detection of aberrant p16 methylation in the plasma and serum of liver cancer patients. *Cancer Research*, *59*(1), 71–73.

Yung, T. K. F., Chan, K. C. A., Mok, T. S. K., Tong, J., To, K. F., & Lo, Y. M. D. (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(6), 2076–2084. <https://doi.org/10.1158/1078-0432.CCR-08-2622>



## Supplementary figures and tables

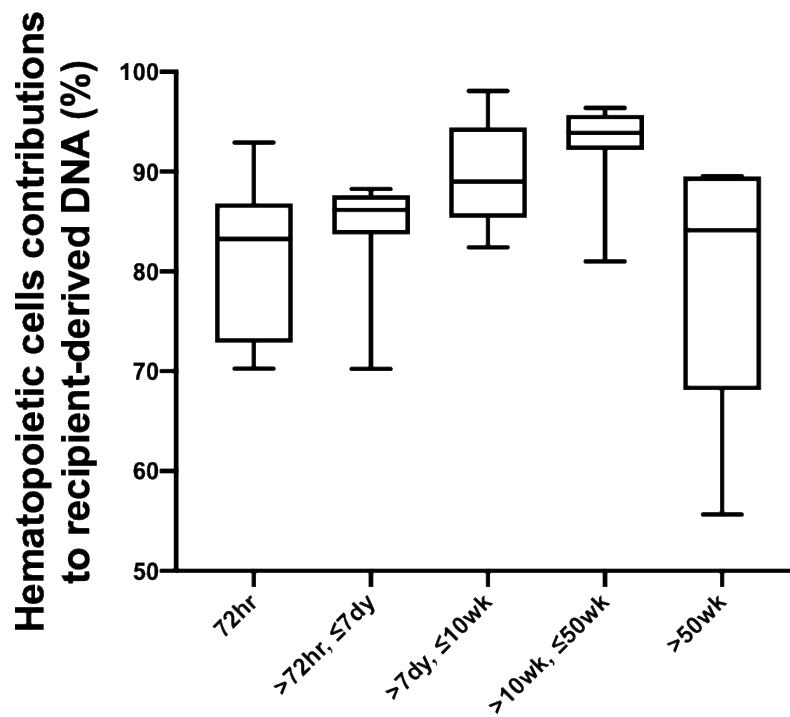


Fig. S1. Percentage contributions of hematopoietic cells to the plasma DNA carrying recipient-specific alleles in patients with lung transplantation.

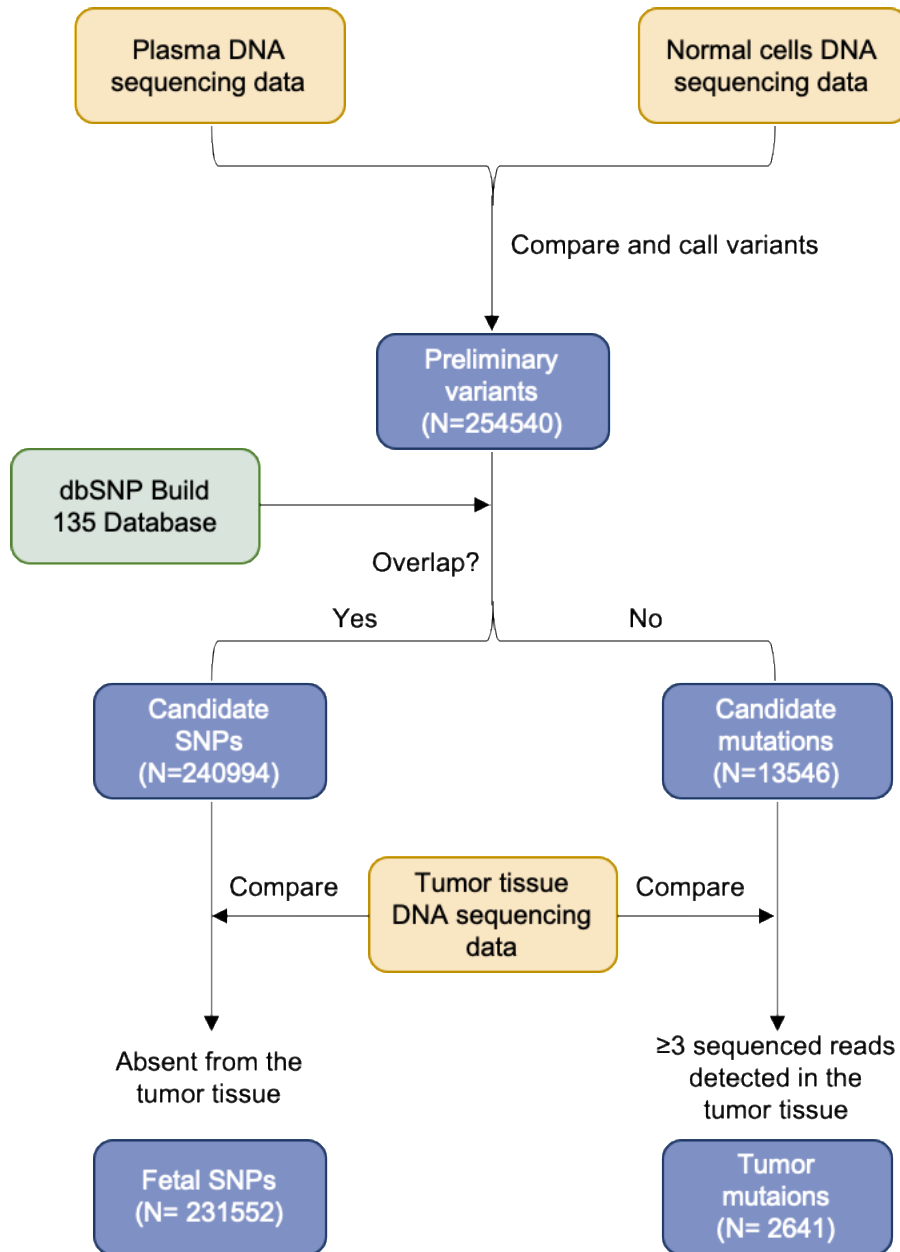


Fig. S2. Flowchart of the steps for identifying the fetal-specific alleles and cancer mutations in the pregnant woman with lymphoma.

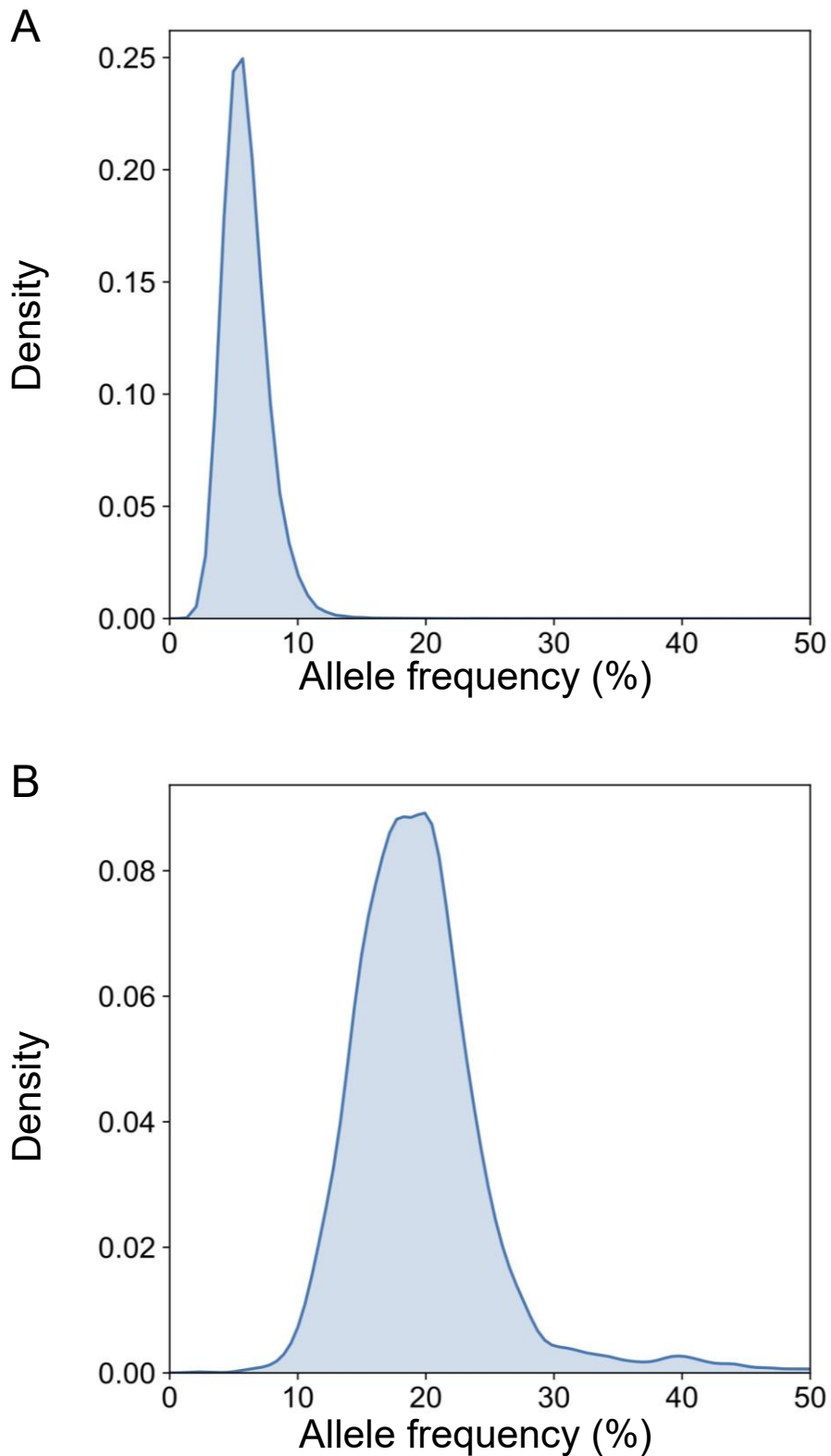


Fig. S3. The distribution of the allele frequency of (A) the fetal-specific alleles and (B) the mutant alleles in the plasma of the pregnant woman with lymphoma.

Table S1. 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 posttransplant
1	34	M	32	M	Cystic fibrosis	double	alive	72hr
2	59	F	27	F	Interstitial lung disease	double	alive	72hr
3	53	M	20	M	Interstitial lung disease	double	alive	72hr
4	63	M	16	F	Interstitial lung disease	double	alive	72hr, 6dy
5	55	F	36	F	Interstitial lung disease	double	alive	72hr, 7dy
6	66	M	48	F	Interstitial lung disease	single	alive	72hr, 4wk
7	66	F	18	M	Chronic obstructive pulmonary disease	single	alive	72hr, 7dy, <u>5wk</u> , <u>20wk</u> , <u>25wk</u> , <u>157wk</u>
8	32	F	39	M	Cystic fibrosis	double	alive	72hr, 7dy, <u>8wk</u> , 38wk, <u>77wk</u> , 129wk
9	67	F	53	F	Sarcoidosis	double	respiratory failure	72hr, 7dy, 6wk, 13wk, <u>22wk</u>
10	44	M	35	F	Re-transplant	double	alive	72hr, 7dy, 10dy, 4wk, 14wk, <u>25wk</u> , 103wk
11	67	F	32	M	Pulmonary arterial hypertension	single	alive	72hr, 7dy, 5wk, 15wk, <u>26wk</u> , <u>61wk</u> , 104wk

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