

RESEARCH ARTICLE

Targeted Sequencing of Genomic Repeat Regions Detects Circulating Cell-free *Echinococcus* DNA

Zhengqing Wan^{1,2}, Xiaoqing Peng^{1,2}, Lu Ma³, Qingshan Tian⁴, Shizheng Wu⁵, Junqi Li⁶, Jie Ling¹, Weigang Lv², Binrong Ding^{1,2}, Jieqiong Tan^{1,2}, Zhuohua Zhang^{1,2,7*}

1 Institute of Molecular Precision Medicine, Xiangya Hospital, Key Laboratory of Molecular Precision Medicine of Hunan Province, Central South University, Changsha, Hunan, China, **2** Center for Medical Genetics, Central South University, Changsha, Hunan, China, **3** School of Biological and Chemical Sciences, Queen Mary University of London, London, United Kingdom, **4** Center for Prevention and Treatment of Echinococcosis, Qinghai Provincial People's Hospital, Xining, Qinghai, China, **5** Department of Neurology, Qinghai Provincial People's Hospital, Xining, Qinghai, China, **6** Sunrain Biotechnology Corporation, Changsha, Hunan, China, **7** Department of Neurosciences, Hengyang Medical School, University of South China, Hengyang, Hunan, China

* zhangzhuohua@skimg.edu.cn



OPEN ACCESS

Citation: Wan Z, Peng X, Ma L, Tian Q, Wu S, Li J, et al. (2020) Targeted Sequencing of Genomic Repeat Regions Detects Circulating Cell-free *Echinococcus* DNA. *PLoS Negl Trop Dis* 14(3): e0008147. <https://doi.org/10.1371/journal.pntd.0008147>

Editor: Klaus Brehm, University of Würzburg, GERMANY

Received: September 4, 2019

Accepted: February 18, 2020

Published: March 10, 2020

Copyright: © 2020 Wan et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All datasets generated or analyzed during this study are included in this published article and its supplementary information files. All in-house scripts used to analyze datasets are available in (https://github.com/wanzhq/Echinococcus_detection).

Funding: This study is supported by a key laboratory grant from Hunan province (2016TP1006), a Science and Technology Major Project of Hunan Provincial Science and

Abstract

Background

Echinococcosis is a chronic zoonosis caused by tapeworms of the genus *Echinococcus*. Treatment of the disease is often expensive and complicated, sometimes requiring extensive surgery. Ultrasonographic imaging is currently the main technique for diagnosis, while immunological analysis provides additional information. Confirmation still needs pathological analysis. However, these diagnostic techniques generally detect infection in late stages of the disease. An accurate, early and non-invasive molecular diagnostic method is still unavailable.

Methodology/Principal findings

We sequenced the cell-free DNA (cfDNA) from plasma of echinococcosis patients and confirmed the presence of *Echinococcus* DNA. To improve detection sensitivity, we developed a method based on targeted next-generation sequencing of repeat regions. Simulation experiments demonstrate that the targeted sequencing is sensitive enough to detect as little as 0.1% of an *Echinococcus* genome in 1 mL of plasma. Results obtained using patient plasma shows that the Area Under the Curve (AUC) of the method is 0.862, with a detection sensitivity of 62.50% and specificity of 100%, corresponding to a Youden-index of 0.625.

Conclusions/Significance

This study provides evidence that hydatid cysts release cfDNA fragments into patient plasma. Using the repeat region targeted sequencing method, highly specific detection of *Echinococcus* infection was achieved. This study paves a new avenue for potential non-invasive screening and diagnosis of echinococcosis.

Technology Department (2018SK1030) and a grant from the National Natural Science Foundation of China (61702555). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: Zhuohua Zhang, Zhengqing Wan and Jie Ling are inventors on a patent application PCT/CN2018/093254 ("Method for Detecting Parasitic Infection and Kit") held by Xiangya Hospital, Central South University that covers the detecting method described in this manuscript. All other authors declare no competing interests.

Author summary

Echinococcosis is a severe chronic parasitic disease caused by tapeworms of the genus *Echinococcus*. According to the World Health Organization, there are more than 1 million people living with echinococcosis worldwide. For decades, little progress has been made to develop a molecular diagnosis and specific treatment for the disease. Although imaging and immunological detection are used for diagnosis, these technologies are either only effective for late stages of the disease or hardly conclusive. The detection of cell-free DNA has been a powerful tool for precise diagnosis. In this study, we showed the presence of *Echinococcus*-derived cell-free DNA in plasma of echinococcosis patients. We further established an assay to detect parasite DNA in blood samples based on amplification of *Echinococcus* specific repeat regions followed by targeted next-generation sequencing. This technique provides a new method for potential extensive screening and precision diagnosis of echinococcosis with high specificity.

Introduction

Echinococcosis is a severe parasitic disease that predominantly affects agricultural and pastoral areas, especially in South America, Africa, and Asia [1, 2]. The etiological agent of echinococcosis are tapeworms of the genus *Echinococcus* [3]. There are mainly two types of the disease spread in the world, cystic echinococcosis (CE) caused by dog-transmitted *E. granulosus* and alveolar echinococcosis (AE) caused by fox-transmitted *E. multilocularis* [2]. Humans are accidental intermediate hosts in the life cycle of *Echinococcus*. CE, also known as hydatid disease, is the most common form all over the world. Once an individual is infected by *E. granulosus*, the larvae of the tapeworm develop into fluid-filled cysts in various organs [4]. As cysts grow larger over a period of time, symptoms caused by compression will arise, and permanent damage to affected organs will occur [2, 5]. In contrast, the larvae of *E. multilocularis* invade in a cancer-like manner to damage the infected organs by infiltrating surrounding tissues, sometimes even metastasizing to other organs.

Clinical symptoms of echinococcosis are atypical and often do not occur until late stages of the disease. Consequently, diagnosis of the disease, especially early diagnosis, is challenging. Ultrasonographic imaging, with the advantage of low cost and rapid diagnostic abilities, is most commonly used [3, 6, 7]. Other imaging techniques with higher resolution such as CT and MRI are employed to detect lesions in specific anatomical locations or atypical echinococcosis [7]. The disadvantage of the imaging techniques is difficult to distinguish *Echinococcus* cysts from other types of cysts. In addition, imaging analysis requires relatively large cysts in CE patients that are already in late stages of the disease [8]. Serological tests may enable earlier diagnosis than imaging technology [9]. Several immunological methods to detect anti-*Echinococcus* antibodies have been developed [9]. However, sensitivity and specificity of these immunological assays vary in different conditions, especially in the cases of CE [10–17]. Therefore, immunological tests are generally used in combination with imaging techniques. Pathological examination of biopsy samples before surgery can be dangerous due to the regeneration capacity of the protoscolex and the risk of anaphylactic reaction during the biopsy procedure [7, 8, 18]. Thus, there is an urgent need for diagnostic methodology that can detect *Echinococcus* infection etiologically and non-invasively.

Assays based on detecting *Echinococcus*-derived circulating antigens were also developed and reported to have high specificity. Unfortunately, their sensitivity was relatively low [17,

19–24]. Consequently, they are barely used in clinic. Likewise, detection of *Echinococcus* DNA by PCR, qPCR or LAMP (loop-mediated isothermal amplification) was also reported [25–31]. These assays are only used for evaluating *Echinococcus* prevalence in dog or fox feces [25, 26, 28, 29, 31] or *Echinococcus* genotyping [27, 30]. Chaya et al. recently reported diagnosis of hydatid disease in humans using PCR detection of parasite DNA in patient serum samples. However, only 25% of hydatid patients appeared positive. All these positive samples had a ruptured cyst confirmed by surgery [32]. No *Echinococcus* DNA was detected in the patient's urine samples [32]. To the best of our knowledge, no reliable molecular detection method is currently used for non-invasive clinical diagnosis of echinococcosis.

Cell-free DNA (cfDNA) refers to extracellular DNA in various body fluids and its detection in secretions such as blood, urine, and saliva, has been received much attention. The main features of cfDNA include fragmentation, low abundance and fast degradation. The size distribution of fragment length of cfDNA is around 166bp that is close to the length of DNA wrapping around a nucleosome [33, 34]. The concentration of cfDNA is 1–10 ng per 1 mL of human plasma under normal circumstances [35]. It increases up to hundreds of nanograms after exercise [36] or under certain disease conditions [37–39]. The half-life of cfDNA in the free state is about 15 min, but becomes more stable when bound to proteins [34, 37, 39]. cfDNA is commonly released during apoptosis and necrosis of cells from different tissues [40, 41]. High-throughput sequencing can be used to analyze the source of cfDNA, therefore, allowing the non-invasive, safe and accurate real-time monitoring of the primary lesions. Also known as “liquid biopsy” [40], cfDNA detection is widely used in prenatal diagnosis [41–43], early tumor detection [44–46] and organ transplant monitoring [47–50]. Methods to detect diseases involving cell death in specific tissues based on cfDNA methylation patterns are also developed [34, 44, 51–53].

Theoretically, DNA fragments of any foreign cells or organisms can be released into host blood, making it possible to be quickly and accurately detected by high-throughput sequencing [54–56]. In the present work, we demonstrate the presence of *Echinococcus*-derived cfDNA in blood plasma of both CE and AE patients. We further established an assay to detect the parasite DNA based on amplification of repeat regions followed by targeted next-generation sequencing (NGS). This technique opens a possibility to extensively screen and diagnose echinococcosis with high specificity and efficiency.

Materials and methods

Sample collection

Plasma and hydatid cyst fluid (HCF) samples from echinococcosis patients were collected from Qinghai province and Xinjiang province in northwest China. All patients fulfilled the diagnostic criteria for echinococcosis [57]. 21 out of 24 patients reported in this study were pathologically confirmed after their blood samples were collected. Three patients (E07, E15 and E20), who had previously diagnosed as echinococcosis and surgically removed cysts, did not take a surgery this time for unwillingness. However, they fitted the diagnostic criteria of Probable cases (E07 and E20) or Possible case (E15) according to WHO-IWGE. CE patients were mainly at stage CE1 or CE2 according to the WHO-IWGE classification. AE patients were at stage P1N0M0 or P2N0M0. For all patients recruited to this study, no chemotherapy was administrated before collecting blood samples. Patients with a sign of either cyst rupture, secondary infection, consolidation or calcification were excluded from the study. Control samples without *Echinococcus* infection and non-relevant control samples with *Schistosoma* infection were collected from Hunan province of central China. The study was approved by the Ethics Committee of Xiangya Hospital of Central South University (Approval No: 201610049).

All participants signed a written informed consent. The Standards for Reporting of Diagnostic Accuracy (STARD) checklist and participants flow chart are provided in [S1 Checklist](#) and [S1 Fig](#), respectively.

For plasma samples, 10 mL whole blood was collected from each patient and control individual using Streck cfDNA blood collection tubes (218997, Streck, USA). The blood was centrifuged at 1600 g for 10 min at 4°C. The plasma was transferred to a fresh 1.5 mL tube, followed by centrifuging at 16,000 g for 10 min at 4°C to remove remaining cell debris. The plasma was used for further cfDNA extraction. For HCF samples, five intact unilocular hydatid cysts were surgically removed from five unrelated cystic echinococcosis patients. After washing with saline, 3~4 mL HCF was carefully aspirated from each cyst. HCF samples were centrifuged at 16,000 g for 10 min at 4°C. Supernatants were transferred to fresh tubes for further cfDNA extraction. Samples comprising 3 mL of plasma or HCF were used to extract cfDNA using the QIAGEN circulating nucleic acid extraction kit (55114, QIAGEN, Germany) according to the manufacturer's protocol.

Untargeted cfDNA sequencing

30 ng cfDNA from plasma or HCF was used to prepare NGS libraries with standard protocol and reagents from NEBNext ultra II library kit (E7645, NEB, USA). Due to short length of cfDNA fragments, no further fragmentation was performed. Libraries were then sequenced by Illumina HiSeq X. 35M~50M reads were generated using the PE150 sequencing strategy. Raw data was first processed using trim-galore (Version 0.4.4) in Paired-end mode with default parameters to cut adapter sequences and remove short (< 20bp) or low-quality (Phred score < 20) reads. Duplicated reads were then removed using FastUniq (Version 1.1). For plasma samples, the de-duplicated reads were first mapped to a human reference genome (version hg38) using the bwa-mem algorithm with default parameters. Unmapped reads were extracted using SAMtools followed by searching against the NCBI nt database that includes 43,107,468 sequences in total (last updated in Jun 2017) using BLAST (Basic Local Alignment Search Tool, command line version 2.6.0+). The identity cut-off was 95%. Reads uniquely assigned to *E. granulosus* or *E. multilocularis* by BLAST were counted. In a "remapping strategy", no-hit reads from the BLAST analysis were collected and mapped to reference genomes of *E. granulosus* (ASM52419v1) or *E. multilocularis* (EMULTI002) using the bwa-mem algorithm with default parameters. For HCF samples, data was preprocessed as described above followed by mapping to a reference genome of *E. granulosus* (ASM52419v1) using bwa-mem. Unmapped reads were extracted and aligned to a human reference genome (version hg38).

Identification of repeat sequences of *E. granulosus* and *E. multilocularis*

Repeat sequences were identified using the RepeatExplorer web server on the Galaxy platform following instruction provided in the on-line manual (<http://repeatexplorer.org/>) [58]. *E. granulosus* and *E. multilocularis* sequencing data (ERR112220 and ERR065034 respectively) downloaded from the SRA databases were used as input sequences. High copy number hits were searched using > 0.01% genome proportion as the criterion. To confirm sequence specificity, each unique high copy number sequence was analyzed using BLAST (command line version 2.6.0+). Repeat sequence hits on species other than *Echinococcus spp.* in the NCBI nt database with a percent of identity > 95% were removed. *Echinococcus* specific repeat sequences with lengths between 70 and 2000bp were selected. These filtered sequences were then used for further primer design.

Primer design and validation

Primers were designed based on identified highly repeated sequences using an on-line primer design tool in NCBI (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>). One or two primer pairs for each repeat sequences were designed with product sizes restricted to 70~100bp and T_m of $60 \pm 1^\circ\text{C}$. Primer specificity was checked in the nr/nt database to exclude those likely to produce unintended products.

Each pair of primers was first validated by conventional PCR with a reaction mixture containing 10 μL of Takara Premix EX Taq (RR030A, Takara, Japan), 1 μL primers with a concentration of 10 $\mu\text{mol/L}$, 100 ng of human genomic DNA with or without 1pg of HCF DNA for the test and control group, respectively. The total reaction volume was brought to 20 μL using ultrapure water (10977023, Invitrogen, USA). The background human DNA was mixed human genomic DNA extracted from whole blood of 5 different individuals. The human genomic DNA was fragmented to size of approximately 160bp using the S2 Focused-ultrasonicator (Covaris, USA). For tests with a reduced amount of template DNA, 100 fg of HCF DNA combined with 10 ng of fragmented human genomic DNA was used as PCR templates. The PCR reaction was performed with an initial denaturation at 95°C for 5 min, followed by 30 cycles of 30 seconds at 95°C for denaturation, 30 seconds at 58°C for annealing and 15 seconds at 72°C for extension. The PCR products of each pair of primers were analyzed on a 2100 bioanalyzer using an Agilent DNA 1000 kit (5067–1504, Agilent, USA). Primer pairs that yielded the intended products were chosen to assemble a multiplex EcDNA (*Echinococcus* cfDNA) primer panel.

The EcDNA primer panel was further verified using HCF DNA samples. The total concentration of the primer panel was 20 $\mu\text{mol/L}$, with an equal molarity of each primer. Multiplex PCR reactions were performed in a total volume of 20 μL containing 2 μL of the primer panel, 1 μL of 10 ng/ μL human genomic DNA, 6 μL of ultrapure water and 1 μL of 1 pg/ μL HCF DNA and 10 μL of Takara Premix EX Taq DNA polymerase. To increase amplification efficiency, 10 additional touch-down cycles were added prior to aforementioned 30 cycles of amplification, by decreasing the annealing temperature 1°C per cycle from 68°C to 58°C . The multiplex PCR products were purified and subjected to NGS sequencing to confirm the sensitivity of each primer pair.

Analysis of patient cfDNA samples

The validated primer panel was used to detect *Echinococcus*-derived DNA in patient plasma cfDNA samples. The multiplex PCR reaction mixtures were composed of 10 μL of Takara Premix EX Taq, 2 μL of the EcDNA multiplex primer panel (20 $\mu\text{mol/L}$), and 8 μL of the circulating cfDNA sample from patients or controls. All samples were renamed before the multiplex PCR procedure for blind evaluation. Reactions were performed using the multiplex PCR conditions with touch down cycles as described above.

The multiplex PCR products were purified using 50 μL of AMPure XP beads (A63881, Beckman Coulter, USA) according to the manufacturer's protocol. To increase recovery efficiency of short products, 5 μL of PEG buffer consisting of 40% PEG8000 and 10 mM EDTA was added to cleanup reactions. For each sample, 10 ng of recovered DNA was used to construct an NGS library using the NEBNext ultra II library prep kit (E7645, NEB, USA) according to the standard protocol. Libraries were then sequenced on an Illumina MiSeq 500 platform (Illumina, USA) in PE75 mode. An average of 2 M reads was generated for each sample.

The raw sequencing data was initially processed using trim-galore as described above. Clean reads were analyzed using an in-house Perl script (<https://github.com/wanzhq/>)

[Echinococcus_detection](#)). Briefly, clean data was initially mapped to the hg19 human reference genome using bowtie2. Unmapped reads were extracted and aligned to each intended PCR products or repeat sequences. The mapping ratio was calculated as the proportion of reads that mapped to *Echinococcus* repeats divided by the total number of reads unmapped to hg19.

Copy number calculation of repeat sequences

The copy numbers of three enriched repeats (Egs-1, Egs-2, Egs-3) were calculated using droplet digital PCR (ddPCR). 5 pg of HCF DNA was used as template for each ddPCR reaction. DdPCRs were done using a Bio-Rad QX200 system (Bio-Rad, USA). A pair of primers specifically amplifying glyoxylate reductase/hydroxypyruvate reductase (EGR-05219) of *E. granulosus* genome (forward: GTGTCTTCAACGACGAGGTTAG; reverse: GTCAGCGTAACCATGCAAATG) was included as control. The copy numbers of repeats were normalized to the genome copies of EGR-05219 in the same reaction.

Statistical analysis

Mapping ratio of unique group were calculated and compared. Data is presented as means \pm SEM. The unpaired t-test with Welch's correction was used to calculate p values. Statistical significance was shown as *p < 0.05, **p < 0.01, ***p < 0.001, and ****p < 0.0001. ROC analysis was performed using SPSS statistical software version 16.0 (IBM Corp., USA).

Results

Detection of *Echinococcus* DNA in plasma of echinococcosis patients and human DNA fragments in hydatid cyst fluids

To examine whether *Echinococcus* DNA is present in patient plasma samples, we initially performed direct NGS analysis of circulating cfDNA from plasma samples of 16 patients and 11 healthy controls. For each sample, sequencing data was first mapped to the human genome (hg38). Subsequently, unmapped reads were extracted followed by searching against the NCBI nt database using the BLAST algorithm. Reads uniquely aligned to *E. granulosus* or *E. multilocularis* were considered as *Echinococcus*-derived sequences. As shown in [Table 1](#), DNA of *E. granulosus* or *E. multilocularis* was detected in 5 out of 16 patient samples, indicating that *Echinococcus* DNA was able to release from hydatid cysts during an infection. Considering the low coverage of the *Echinococcus* genome sequences in the nt database, we collected reads that did not generate hits in the BLAST analysis followed by remapping these reads to the genome of *E. granulosus* (ASM52419v1) or *E. multilocularis* (EMULTI002) using bwa-mem algorithm. These remapped reads can be regarded as a collection of sequences most likely derived from the *Echinococcus* genome. The ratio of remapped reads in patients was significantly higher than that in controls ([S2 Fig](#)). *Echinococcus* DNA was detected in 9 out of 16 patients with remapping ratio above 100×10^{-6} ([Table 1](#)). For AE, parasite-derived DNA was detected in 4 out of 5 patients.

To further verify the wall of hydatid cysts could allow DNA fragments to pass through, we extracted cfDNA from the HCF of intact hydatid cysts from 5 unrelated patients followed by NGS. Despite limited contamination with large fragments of genomic DNA, predominant size of cfDNA from HCF was approximately the same as that of human plasma cfDNA ([Fig 1A](#)). Sequencing reads mapped to the human reference genome were detected in all 5 HCF samples ([Table 2](#)). The proportion of human-derived DNA sequences reached as high as 80% of the total cfDNA sequencing reads. The lowest ratio of reads within one cyst that matched to the human genome was about 12%. Results suggest that the hydatid cyst wall is permeable at least

Table 1. BLAST analysis and realignment of plasma cfDNA of echinococcosis patients and controls.

Group	Samples	Mapped to Human Genome	Reads Unmapped to Human Genome	Unique BLAST hits for <i>E. g</i> or <i>E. m</i>	Reads without BLAST hits	Reads Mapped to <i>E. g</i> or <i>E. m</i>	Ratio to <i>E. g</i> or <i>E. m</i> ($\times 10^{-6}$)
Cystic Echinococcosis Patients	E01	98.78%	666337	0	362558	31	85.504
	E02	98.14%	987872	0	584146	80	136.952
	E03	97.82%	789692	0	461545	261	565.492
	E05	98.14%	884296	0	511312	78	152.549
	E06	98.92%	461088	0	302465	6	19.837
	E09	98.57%	592295	0	448111	8	17.853
	E16	98.89%	1698688	0	1417945	91	64.177
	E17	99.55%	138359	0	117844	0	0.000
	E18	99.31%	203737	2	158564	54	340.556
	E19	99.73%	186700	0	94623	29	306.479
Alveolar Echinococcosis Patients	E20	98.65%	507440	0	373063	9	24.125
	E07	99.30%	389896	2	290186	205	706.443
	E08	98.98%	456714	6	346458	101	291.522
	E10	99.23%	382636	0	328750	12	36.502
	E11	99.20%	367624	32	259077	1173	4527.611
Normal Controls	E12	99.27%	344414	21	248630	698	2807.384
	CT01	99.31%	473305	0	327226	4	12.224
	CT02	99.24%	431979	0	265245	2	7.540
	CT03	98.59%	884409	0	813749	23	28.264
	CT04	99.30%	212610	0	186636	6	32.148
	CT05	99.35%	189895	0	171294	3	17.514
	CT06	98.65%	708331	0	378457	24	63.415
Schistosomiasis Patients	CT07	98.68%	618738	0	504094	29	57.529
	S01	99.32%	224216	0	180962	5	27.630
	S02	99.43%	215004	0	176148	5	28.385
	S03	99.56%	154415	0	133285	2	15.005
	S04	99.46%	154279	0	129377	0	0.000

E. g.: *E. granulosus*. *E. m.*: *E. multilocularis*. BLAST: NCBI Basic Local Alignment Search Tool. Unique BLAST hits for *E. g* or *E. m*: number of reads that were only matched to *E. g* or *E. m* sequences in the NCBI nt database. Reads without BLAST hits: number of reads that did not match any sequences in the NCBI nt database. Reads Mapped to *E. g* or *E. m*: number of reads that did not generate BLAST hits but could be mapped to the genome of *E. g* or *E. m* using bwa-mem algorithm. Ratio to *E. g* or *E. m*: mapping ratio of the mapped reads set divided by the set of reads without BLAST hits.

<https://doi.org/10.1371/journal.pntd.0008147.t001>

to short, freely circulating nucleic acids, such as cfDNA. In these cases, further mapping analysis revealed that most of the identified sequences of *E. granulosus* cfDNA were from non-coding, low complexity regions. In contrast, human cfDNA sequences were from both coding and non-coding regions of the genome (Fig 1B).

Multiplex PCR panel design

To increase detection sensitivity of cfDNA at trace levels, we employed a targeted multiplex PCR of repeat regions to enrich *Echinococcus*-derived cfDNA before NGS. A total of 19732 repeat sequences were identified using RepeatExplore on the Galaxy platform (S1 Appendix), based on *E. granulosus* (ERR112220) and *E. multilocularis* (ERR065034) sequences downloaded from the NCBI SRA database. To improve the template utilization efficiency, length of products was restricted to 70~100bp considering ~166bp of typical cfDNA. A total of 16 primer pairs were initially selected for validation (Table 3).

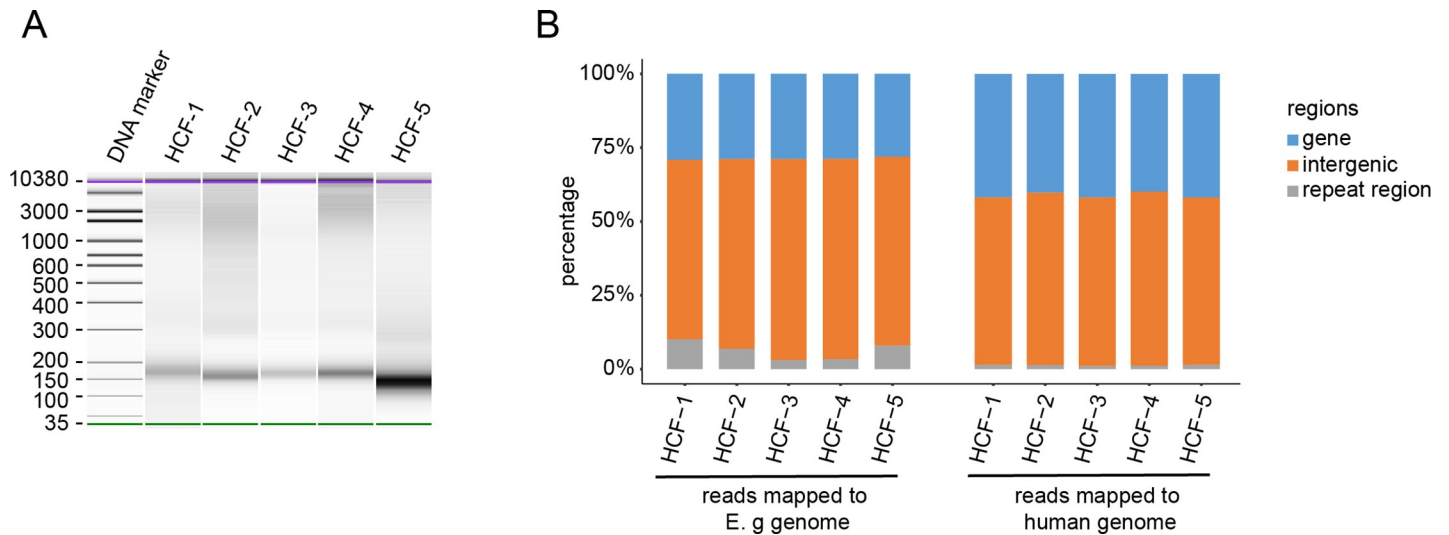


Fig 1. Profiles of cfDNA extracted from intact hydatid cysts. (A) Size of cfDNA from HCF samples analyzed on a 2100 bioanalyzer. Molecular weight (bp) is indicated (left lane). Green lines:15bp. Purple lines: 1500bp. (B) Profile of the cfDNA sequencing reads mapped to the *E. granulosus* genome (ASM52419v1) and human genome (hg38).

<https://doi.org/10.1371/journal.pntd.0008147.g001>

Each pair of candidate primers were validated by conventional PCR using a mixture of HCF-1 HCF cfDNA and human genomic DNA as templates. A total of 12 pairs of primers generated anticipated products with 1 pg HCF cfDNA mixed with 100ng of human genomic DNA. On the other hand, no specific band was detected using human DNA alone as templates (Fig 2A–2C). Further analysis revealed that 10 pairs of primers were sensitive enough to obtain specific PCR products using 100 fg of HCF cfDNA as templates (Fig 2D and 2E). 13 pairs of primers yielded specific PCR products using DNA extracted from the liver lesion of an alveolar echinococcosis patient (S3 Fig). As our primary goal was to detect CE, 12 pairs of primers that gave specific products with HCF DNA were combined into a EcdNA (*Echinococcus* cfDNA) primer panel. Multiplex PCR followed by NGS showed that the primer panel is effective to amplify *Echinococcus* DNA from 1 pg HCF DNA (Fig 3A, Table 4). Considering that human DNA is present in hydatid cysts, we further analyzed the copy number of *Echinococcus* genome per 100 pg HCF DNA by droplet digital PCR. The results showed that the copy number of the *Echinococcus* genome per 100 pg HCF DNA varied from 20 to 400, which was consistent with results of NGS analysis (Table 2 and S4 Fig). The approximate repeat numbers for Egs-1, Egs-2 and Egs-3 per one *Echinococcus* genome were 7000, 800 and 400, respectively (S4 Fig).

NGS analysis of multiplex PCR products revealed that the ratio of reads mapped to repeat sequence dataset (98.93% on average) was higher than that of reads exactly mapped to the

Table 2. Alignment results of the reads obtained from HCF sequencing.

Samples	Total Reads	Reads Mapped to the E. g Genome	Percentage of Reads Mapped to E. g genome	Reads Mapped to the Human Genome	Percentage of Reads Mapped to Human genome
HCF-1	50975739	7216672	14.16%	43343421	85.03%
HCF-2	15738707	3723808	23.66%	11924502	75.77%
HCF-3	15827757	12114428	76.54%	2883866	18.22%
HCF-4	15325904	12961579	84.57%	1863580	12.16%
HCF-5	33656866	6281636	18.66%	27331719	81.21%

HCF: hydatid cyst fluid; E. g: *E. granulosus*.

<https://doi.org/10.1371/journal.pntd.0008147.t002>

Table 3. Primers selected and their target repeat sequences.

Primer Pairs	Repeat Names	Forward Primer	Reverse Primer	Minimal Product Length (bp)**
Egs-1	>CL11Contig2	CACTGTGACGTCATCTGGCCT	TCAGGTGACGTAATGGAGGGTCT	70
Egs-2	>CL22Contig1	TCTGTGCTTGCATTACAC	CAAATGCTCGGTACACCACG	84
Egs-3	>CL34Contig1	CTGCAATAGCACCCAATTCACA	GAAGGAGTATCGTTGGTACGCT	86
mgs-1	>CL5Contig3	TTCGTGCTACGACTTCTCCAC	GGAGTGCAAATGAAGTAGATGCG	72
mgs-2	>CL10Contig9	TCAAGTATGTTGCGAAGGCGA	TGCATGGTAGAGACCCGGAA	76
mgs-3	>CL13Contig14	GTTGCCAGGGCAGTGAGTTA	TGGCATTGGGGCGTGAAGTAG	71
mgs-4	>CL17Contig31	AGTAGCGGAACGGTGGATT	ACAATGGCCGGTAGTGAAGA	83
mgs-5	>CL19Contig16	CCACCCAGCGAGGTACAAG	AGTGGTTATCCCTCGGTTCTG	75
mgs-6	>CL21Contig4	GTTTCACACCGACAACAGTCAA	TGAGTCGAAGGCGAACACC	76
mgs-7	>CL23Contig4	GGATCCGTCGATGTTAGCGT	CCGCCATAGAGGGTAGATGC	71
mgs-8	>CL10Contig9	ATCCGCACCTGTCGTAACCT	GTAACATGCGGCTTCGGAAT	88
mgs-9	>CL13Contig14	CGGGACACATCCTAACTGGT	CCGGTCATCCATGGGGATTG	83
mgs-10	>CL23Contig4	CAGAGGCTCGTTTGTGGTCA	GGTGCACATTAAATACCAAACCC	73
mgs-11	>CL26Contig5	CGAAGGACAGCCATTTCCGA	TCAGCGAGCCACAGATTACAT	70
mgs-12	>CL27Contig3	TGGCGCAACACCTTGTAGAT	GAAGGTGAAGGTGCCGAAGA	87
mgs-13	>CL29Contig14	homology to human histone gene*		/
mgs-14	>CL30Contig2	AGCACTCCTCATCAGTCAACTC	CTGAAACATGCTAAAGGTATGCGT	82

* Repeat sequence >CL29Contig14 was excluded because it is highly homologous to human histone genes.

** Size of actual products may be different from predicted due to the possibility of tandem repeat and indels in repeats.

<https://doi.org/10.1371/journal.pntd.0008147.t003>

anticipated PCR products (91.75% on average). The difference likely resulted from the variability of each repeat in genome (Fig 3B, Table 4). Therefore, further analysis was based on the ratio of reads mapped to the repeat regions rather than those mapped to the anticipated PCR products.

The EcDNA primer panel detect *Echinococcus* DNA using femtogram HCF cfDNA as templates

To further determine the detection limit of the EcDNA primer panel, we titrated HCF cfDNA templates mixed with human cfDNA. The results showed that the EcDNA panel was sensitive enough to stably detect as little as 2 fg cfDNA from the HCF-1 sample (Fig 4, S1 Table). To further verify the detection sensitivity, we added various quantities of HCF cfDNA to plasma isolated from healthy controls, followed by extracting cfDNA from the mixed plasma. cfDNA isolated from the mixtures were subjected to multiplex PCR and NGS. The results revealed that the detection sensitivity was as little as 5 fg HCF cfDNA per 1 mL of plasma (S2 Table). The quantitative results of ddPCR suggest approximately 30 copies of *Echinococcus* genome per 100 pg of DNA in the HCF-1 sample (S4 Fig). Therefore, the method is sensitive enough to detect 0.1% of *Echinococcus* genome per 1mL of plasma.

Detection of *Echinococcus* DNA using cfDNA from patient plasma samples

The EcDNA primer panel was next used to analyze cfDNA from plasma samples of patients infected with *Echinococcus* (n = 24). Healthy uninfected individuals (n = 27) and patients infected with *Schistosoma* (n = 9) were also analyzed as controls. A clinical summary of controls and patients is shown in Table 5. Since no method is available to exclude *Echinococcus* infection in early stages, all control samples were collected from Hunan province of central China, where echinococcosis is rarely reported. Further follow-up of the controls finds no

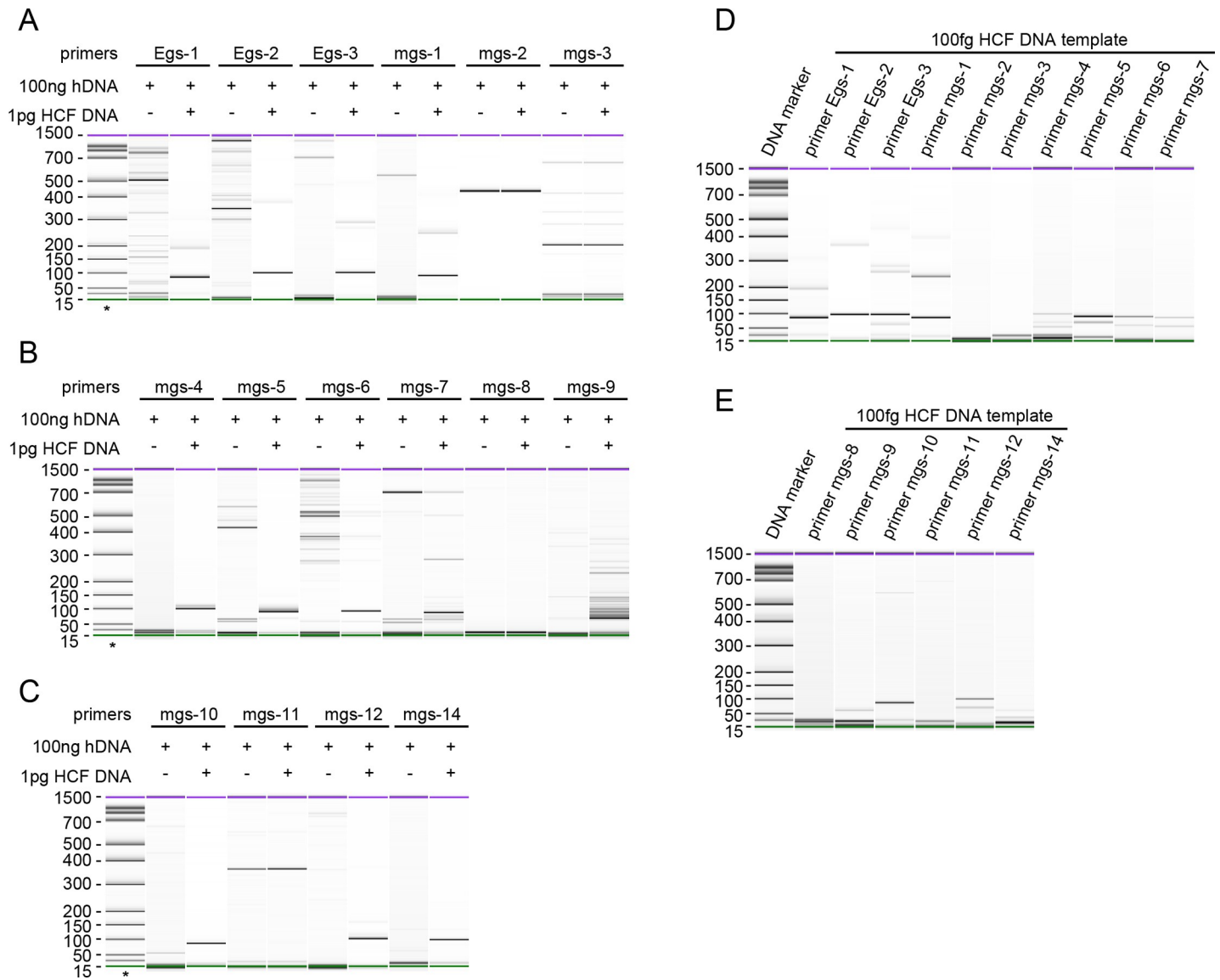


Fig 2. Validation of primers via conventional PCRs. (A, B, C) Conventional PCR amplification with 1 pg of HCF DNA as template in the background of 100 ng of human genomic DNA with each primer pair. Expected products were observed with the primer pairs Egs-1, Egs-2, Egs-3, mgs-1, mgs-4, mgs-5, mgs-6, mgs-7, mgs-9, mgs-10, mgs-12 and mgs-14. (D, E) PCR sensitivity analysis by decreasing the HCF DNA template to 100 fg. Products of the expected size were observed with 10 primer pairs. The molecular weight (bp) is indicated on the left of each panel. * indicates DNA marker. Green lines:15bp, Purple lines: 1500bp.

<https://doi.org/10.1371/journal.pntd.0008147.g002>

infection with *Echinococcus* of these individuals. Detailed information of patients is shown in Table 6.

The results showed that the mapping ratio, calculated as the proportion of reads mapped to the *Echinococcus* DNA repeats over the total number of reads unmapped to the human reference genome, was significantly higher in samples from alveolar and cystic echinococcosis patients than those of schistosomiasis patients or uninfected control individuals (Fig 5). *Echinococcus*-derived cfDNA was detected in 15 out of 24 plasma samples from echinococcosis patients with cut-off to 10.0%. In contrast, no *Echinococcus*-derived DNA was detected in the cfDNA isolated from plasma of schistosomiasis patients and healthy control individuals (Table 7). Results indicate that the targeted detection method is more sensitive than the direct sequencing method.

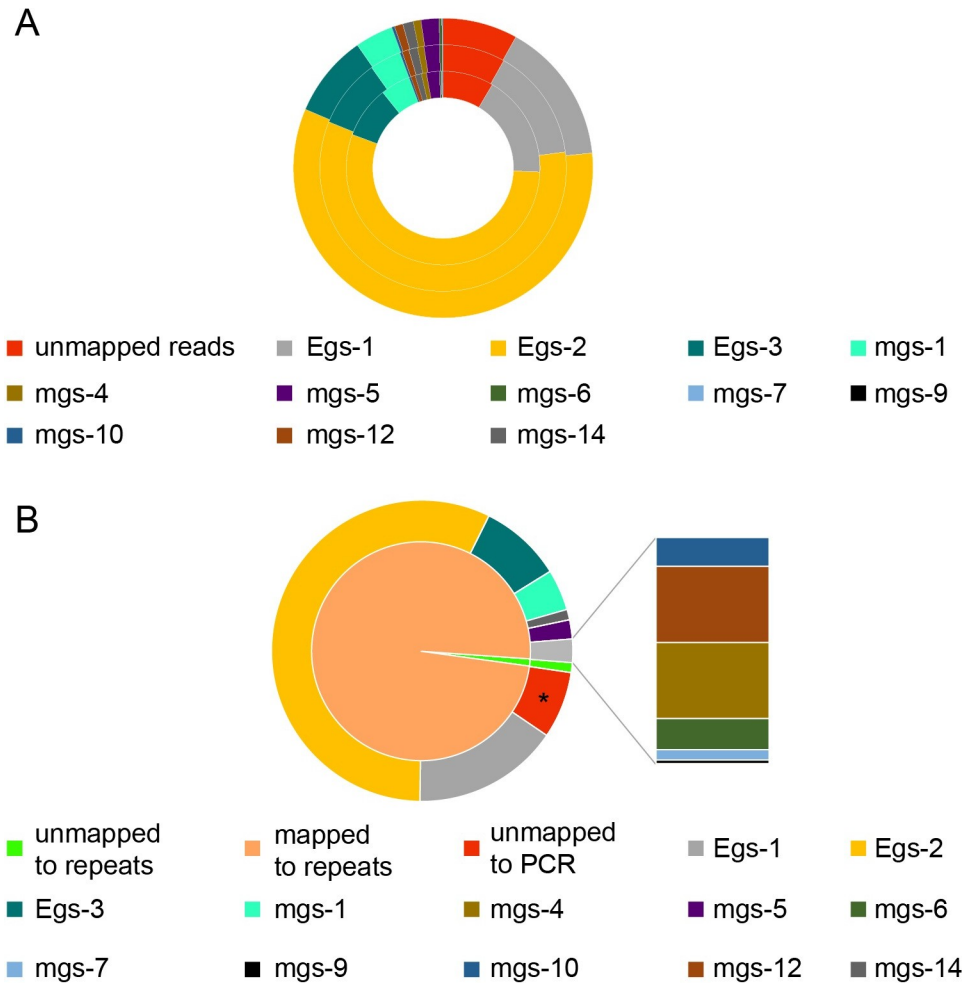


Fig 3. PCR products observed in multiplex PCR. (A) Results of three independent samples are shown. Each pair of primers yielded a product of expected sequence. (B) Comparison of average mapping ratio to repeat sequences and intended products. * Part of the reads unmapped to the intended products were mapped to repeat sequences.

<https://doi.org/10.1371/journal.pntd.0008147.g003>

A ROC analysis based on the mapping ratio of plasma samples showed that the area under the curve (AUC) was 0.862, with a 95% confidence interval of 0.758~0.965 (Fig 6). When the cut-off was set to 9.79%, the ROC curve reached a Youden-index of 0.625, with a detection sensitivity of 62.5% and specificity of 100%.

Discussion

Echinococcus parasitizes deeply in human organs such as liver, lung and brain. Due to the risk of disseminated infection, it is not advisable to conduct a biopsy to make a definitive diagnosis. Thus, non-invasive methodology for early diagnosis is urgently needed. In theory, hydatid cysts allow nutrients in and metabolites out in order to grow [59]. In this study, we show that *Echinococcus* derived cfDNA is present in plasma of *Echinococcus* infected individuals, which could facilitate non-invasive diagnosis. We further designed a targeted NGS strategy to detect *Echinococcus*-derived cfDNA in blood samples. The results suggest that targeted NGS is feasible to achieve non-invasive molecular diagnosis of echinococcosis.

Targeted NGS detected *Echinococcus* cfDNA in patient plasma in this study. The finding provides the first evidence to our knowledge that *E. granulosus* DNA is also released from

Table 4. Validation of the EcDNA primer panel using multiplex PCR followed by NGS.

	1pg-1	1pg-2	1pg-3	Average
Hit for Repeats	98.32%	99.26%	99.22%	98.93%
Hit for PCR	91.51%	91.81%	91.93%	91.75%
Mapped to Each Primer Product				
Egs-1	17.17%	14.74%	15.34%	15.75%
Egs-2	54.98%	58.20%	58.02%	57.06%
Egs-3	8.64%	9.04%	8.88%	8.85%
mgs-1	4.98%	4.13%	4.12%	4.41%
mgs-4	0.86%	0.85%	0.83%	0.85%
mgs-5	2.17%	2.03%	1.94%	2.05%
mgs-6	0.34%	0.35%	0.32%	0.33%
mgs-7	0.21%	0.08%	0.08%	0.12%
mgs-9	0.03%	0.04%	0.04%	0.04%
mgs-10	0.32%	0.31%	0.31%	0.31%
mgs-12	0.73%	0.91%	0.91%	0.85%
mgs-14	1.07%	1.15%	1.14%	1.12%

Hit for Repeats: mapping ratio to the repeat sequences. Hit for PCR: mapping ratio to intended PCR products. 1pg-1, 1pg-2, 1pg-3: three independent experiments using 1pg HCF DNA as multiplex PCR templates.

<https://doi.org/10.1371/journal.pntd.0008147.t004>

intact hydatid cysts into blood of infected individuals, laying the basis for a non-invasive and precise diagnosis of echinococcosis using DNA sequencing. It is likely that the identified *Echinococcus* DNA fragments pass through the wall of hydatid cysts to enter the blood stream. This notion is further supported by the discovery of human DNA fragments inside hydatid cysts, despite that patient cfDNA in hydatid cysts is remarkably higher than parasite cfDNA in patient serum.

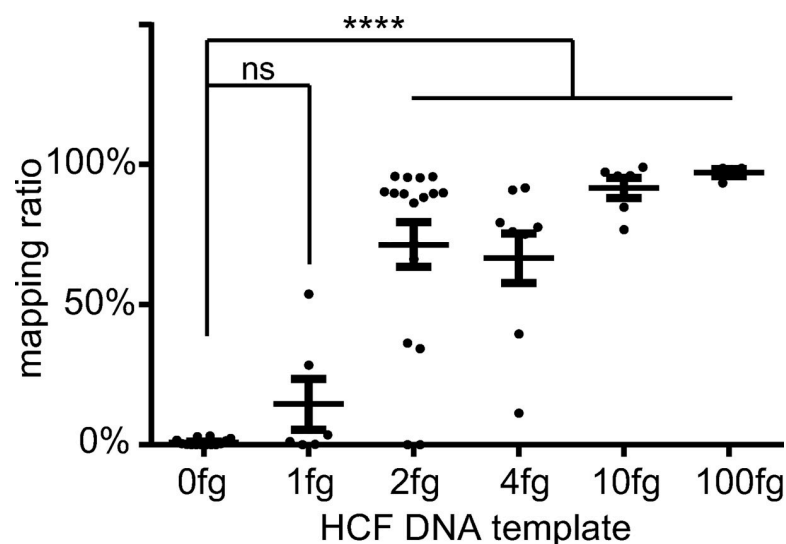


Fig 4. Simulation experiments with various quantities of HCF DNA template. The mapping ratio to the *Echinococcus* genome indicates the ratio of reads mapped to repeat sequences over the total number of reads. Data are expressed as means \pm SEM. **** $p < 0.0001$, unpaired t-test with Welch's correction. Note that detection limit of the EcDNA (*Echinococcus* cfDNA) primer panel was as low as 2 fg/reaction.

<https://doi.org/10.1371/journal.pntd.0008147.g004>

Table 5. Overall information of plasma samples.

Sample Group	Echinococcosis patients (n = 24)	Schistosomiasis patients (n = 9)	Uninfected controls (n = 27)
Age (years)	38.08 ± 2.828	42.78 ± 3.609	36.60 ± 2.261
Gender (male/female)	11/13	5/4	14/13
Imaging-based Examination (positive/negative/not determined)	24/0/0	0/9/0	0/11/16
Immunological Examination (positive/negative/not determined)	18/2/4	0/0/9	0/0/27
Pathological Examination (positive/negative/not determined)	21/0/3	0/0/9	0/0/27

The age was shown as mean ± SEM.

<https://doi.org/10.1371/journal.pntd.0008147.t005>

The concentration of *Echinococcus* cfDNA in patient plasma is likely very low, suggesting that release of *Echinococcus* DNA fragments from hydatid cysts is either limited or that the DNA is rapidly degraded. It is possible that protein-bounded DNA has a limited capacity to pass through the wall of hydatid cysts. The proportion of *Echinococcus* DNA in the cfDNA isolated from patient plasma varies between 1/1000 and 1/100000, which is a major barrier for clinical detection using untargeted NGS. We therefore developed a targeted multiplex PCR primer panel to amplify repeat sequences from the *Echinococcus* genome in order to specifically enrich *Echinococcus*-derived cfDNA for further analysis via NGS. The combination of multiplex PCR and NGS incorporates both the sensitivity of PCR and the specificity of DNA sequencing while avoiding cross-reactions or false positives. This method is sensitive enough to detect a thousandth of an *Echinococcus* genome in simulation experiments. Meanwhile, targeted NGS cut down the sequencing cost from hundreds of dollars for an untargeted sequencing experiment to less than 1 dollar for a targeted sequencing detection. The overall cost for each detection is approximately \$50 in laboratory that can be reduced to less than \$20 with further improvement.

Several PCR or qPCR methods for the detection of *Echinococcus* DNA in patients have been reported by various research groups [25–31]. However, these methods have not demonstrated non-invasive diagnostic values for echinococcosis. Potential reasons could be diverse. One possibility is the low abundance of *Echinococcus* DNA in patient plasma. Alternatively, DNA extraction methods used in previous studies are generally optimized for genomic DNA extraction that is not optimized for short cfDNA fragments. Furthermore, the designed length of PCR products was usually longer than the average length of cfDNA fragments (166bp), reducing detection of short DNA fragments. A recent study on the detection of parasite-derived cfDNA from serum of alveolar echinococcosis (AE) patients partly improved cfDNA extraction and primer design protocol. These authors were able to distinguish all samples from an AE animal model and 30% of patient serum samples via qPCR and ddPCR [60]. The present study enriched targeted DNA fragments by multiplex PCR followed by NGS, which increased the possibility to detect low abundant and short cfDNA fragments. In the present study, the AUC was 0.862, with a detection sensitivity of 62.50% and specificity of 100%, corresponding to a Youden-index of 0.625. Interestingly, the sensitivity increased to 80% if we consider only the 5 alveolar echinococcosis patients. This is likely because of its distinct way of expansion of *E. multilocularis* in host body allowing cfDNA to release to blood stream. The detection sensitivity may be even higher for early infected patients before formation of a complete cyst, because the laminated layer serves as a physiochemical barrier that reduces the permeability of macromolecules [5]. Unfortunately, no technology that diagnose *Echinococcus* infection in very early stages is available. A more comprehensive clinical study with larger numbers of patients will be needed to further verify the effectiveness of the method for early diagnosis.

Table 6. Detailed clinical information of echinococcosis patients.

Sample ID	Gender	Age (years)	Nationality	Clinical Diagnosis	Previous <i>Echinococcus</i> Infection History*	Imaging-based Examination	Immunological Examination	Stages**	Specific Signs	Pathological Diagnosis
E01	male	36	Chinese-Tibetan	hepatic cystic echinococcosis	N	P	P	CE1	N	cystic echinococcosis
E02	male	41	Chinese-Han	diaphragmatic cystic echinococcosis	N	P	N	CE3	N	cystic echinococcosis
E03	male	43	Chinese-Tibetan	pelvic and abdominal cystic echinococcosis	N	P	P	ND	N	cystic echinococcosis
E05	female	43	Chinese-Tibetan	diaphragmatic cystic echinococcosis	N	P	P	CE2	N	cystic echinococcosis
E06	female	34	Chinese-Tibetan	abdominal cystic echinococcosis	P	P	P	ND	N	cystic echinococcosis
E07	female	34	Chinese-Tibetan	hepatic alveolar echinococcosis	P	P	P	P2N0M0	N	ND
E08	female	36	Chinese-Tibetan	hepatic alveolar echinococcosis	P	P	P	P1N0M0	N	alveolar echinococcosis
E09	female	29	Chinese-Han	hepatic cystic echinococcosis	P	P	ND	CE1	N	cystic echinococcosis
E10	female	36	Chinese-Tibetan	hepatic alveolar echinococcosis	N	P	P	P1N0M0	N	alveolar echinococcosis
E11	female	20	Chinese-Tibetan	hepatic alveolar echinococcosis	P	P	P	P2N0M0	left hepatic vein, hepatic segment inferior vena cava involvement	alveolar echinococcosis
E12	female	40	Chinese-Tibetan	hepatic alveolar echinococcosis	P	P	P	P2N0M0	right hepatic vein, right portal vein, inferior vena cava involvement	alveolar echinococcosis
E14	male	41	Chinese-Tujia	hepatic cystic echinococcosis	N	P	P	CE2	N	cystic echinococcosis
E15	female	68	Chinese-Tibetan	hepatic cystic echinococcosis	P	P	N	CE2	residual cavity infection	ND
E16	male	27	Chinese-Uighur	hepatic cystic echinococcosis	N	P	ND	CE2	N	cystic echinococcosis
E17	female	37	Chinese-Kazak	hepatic cystic echinococcosis	N	P	ND	CE1	N	cystic echinococcosis
E18	female	30	Chinese-Hui	hepatic cystic echinococcosis	N	P	ND	CE1	N	cystic echinococcosis
E19	male	7	Chinese-Uighur	hepatic cystic echinococcosis	N	P	P	CE1	N	cystic echinococcosis
E20	male	45	Chinese-Uighur	hepatic and abdominal cystic echinococcosis	P	P	P	CE2	sinus formation	ND
E21	male	27	Chinese-Uighur	hepatic cystic echinococcosis	N	P	P	CE1	N	cystic echinococcosis
E22	male	70	Chinese-Han	hepatic cystic echinococcosis	N	P	P	CE1	multiple hydatid cyst	cystic echinococcosis
E23	female	38	Chinese-Kazak	hepatic cystic echinococcosis	N	P	P	CE2	N	cystic echinococcosis
E24	male	63	Chinese-Han	hepatic cystic echinococcosis	N	P	P	CE1	N	cystic echinococcosis
E25	male	33	Chinese-Hui	hepatic cystic echinococcosis	N	P	P	CE2	N	cystic echinococcosis
E26	female	36	Chinese-Uighur	hepatic cystic echinococcosis	N	P	P	CE1	N	cystic echinococcosis

P: positive. N: negative. ND: not determined.

* Infection history with the same type of echinococcosis.

**Stages were determined according to the WHO-IWGE classification.

<https://doi.org/10.1371/journal.pntd.0008147.t006>

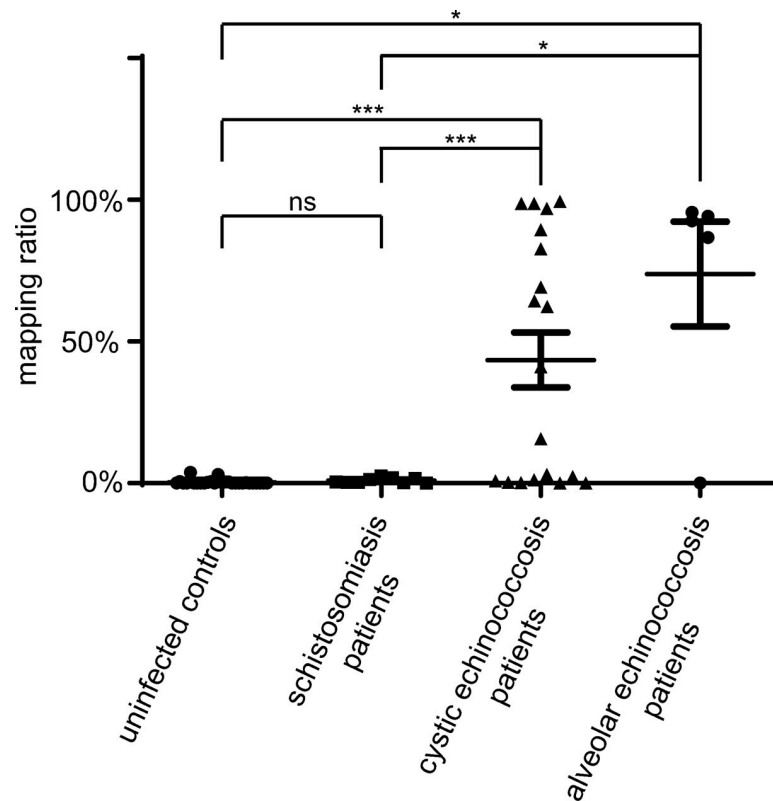


Fig 5. The mapping ratio to the *Echinococcus* genome of plasma cfDNA samples from CE, AE patients and control individuals. Data are expressed as means ± SEM. *** p < 0.001, * p < 0.05, unpaired t-test with Welch's correction.

<https://doi.org/10.1371/journal.pntd.0008147.g005>

Table 7. Mapping ratio to *Echinococcus* repeat sequences of circulating cfDNA of echinococcosis patients and controls.

Cystic echinococcosis Patients	Mapping Ratio	Alveolar echinococcosis Patients	Mapping Ratio	Uninfected Controls	Mapping Ratio	Schistosomiasis Patients	Mapping Ratio
E01	82.74%	E07	92.56%	CT01	0.52%	S01	0.38%
E02	15.74%	E08	86.66%	CT02	0.03%	S02	1.88%
E03	96.94%	E10	0.01%	CT03	3.84%	S03	0.34%
E05	0.22%	E11	95.62%	CT04	0.38%	S04	0.30%
E06	89.45%	E12	94.13%	CT05	0.06%	S05	1.26%
E09	0.76%			CT06	3.08%	S06	2.51%
E14	41.03%			CT07	0.21%	S07	1.74%
E15	98.67%			CT08	0.32%	S08	0.13%
E16	98.73%			CT09	0.55%	S09	0.02%
E17	0.02%			CT10	0.00%		
E18	99.49%			CT11	0.01%		
E19	69.15%			CT12	0.11%		
E20	62.31%			CT13	0.01%		
E21	2.34%			CT14	0.00%		
E22	64.31%			CT15	0.01%		
E23	1.14%			CT16	0.01%		
E24	3.06%			CT17	0.29%		

(Continued)

Table 7. (Continued)

Cystic echinococcosis Patients	Mapping Ratio	Alveolar echinococcosis Patients	Mapping Ratio	Uninfected Controls	Mapping Ratio	Schistosomiasis Patients	Mapping Ratio
E25	0.32%			CT18	0.45%		
E26	0.04%			CT19	0.01%		
				CT20	0.01%		
				CT21	0.01%		
				CT22	0.01%		
				CT23	0.02%		
				CT24	0.03%		
				CT25	0.02%		
				CT26	0.09%		
				CT27	0.04%		

Mapping ratio was calculated as the proportion of reads mapped to the *Echinococcus* DNA repeats over the total number of reads unmapped to the human reference genome.

<https://doi.org/10.1371/journal.pntd.0008147.t007>

Our data provides evidence that hydatid cysts release cfDNA into patient plasma even if the cyst membrane remains intact. This study also establishes a new and practical method based on the concept of cell-free DNA in conjunction with the high sensitivity of multiplex PCR and the specificity of NGS, paving a new avenue for a potential early, accurate and non-invasive diagnosis of echinococcosis.

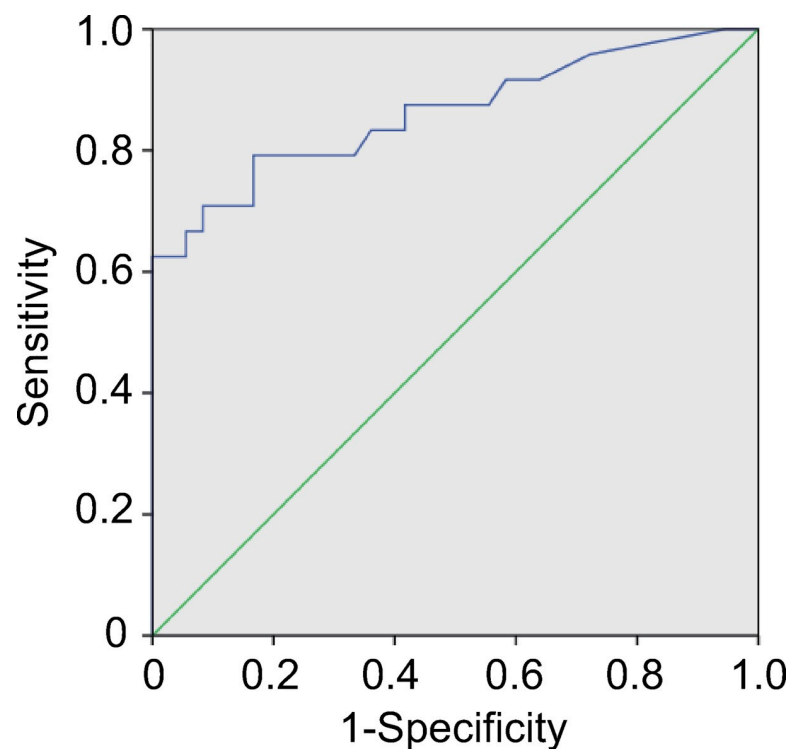


Fig 6. ROC curve based on the mapping ratio of repeat regions with echinococcosis patient samples. Blue line: the ROC curve based on the results of targeted sequencing of repeat regions in this study. Green line: random classification.

<https://doi.org/10.1371/journal.pntd.0008147.g006>

Supporting information

S1 Checklist. Standards for Reporting of Diagnostic Accuracy (STARD) checklist.
(DOCX)

S1 Fig. Flow diagram of participants enrolled in this study.
(TIF)

S2 Fig. Remapping ratios of plasma cfDNA from echinococcosis patients and control individuals. Remapping ratio was transformed by logarithm. E17 and S04 were discarded as outliers because their remapping ratio was zero. The data are expressed as mean \pm SEM. * $p < 0.05$, ** $p < 0.01$, unpaired t-test with Welch's correction.
(TIF)

S3 Fig. Test of primer pairs with *E. multilocularis* DNA. PCR amplification of 100 pg genomic DNA extracted from liver lesion of an alveolar echinococcosis patient in the background of 10 ng human genomic DNA with each pairs of primers. Products of the expected size were observed with primers Egs-2, mgs-1, mgs-2, mgs-3, mgs-4, mgs-5, mgs-6, mgs-7, mgs-8, mgs-9, mgs-11, mgs-12 and mgs-14. The molecular weight (bp) is indicated on the left of each panel. Green lines:15bp. Purple lines: 1500bp.
(TIF)

S4 Fig. Copy number evaluation of *E. granulosus* genome and repeat. (A) Copy numbers per 100 pg of *E. granulosus* genome from HCF samples are shown. (B) Copy numbers amplified by indicated 3 pair of primers in one *Echinococcus* genome. Data is expressed as mean \pm SD.
(TIF)

S1 Table. Mapping ratio of sequencing reads from simulation experiments using various quantities of HCF templates.
(XLSX)

S2 Table. Alignment results of plasma samples added HCF DNA. Indicated quantity of HCF DNA was added per 1mL of plasma from health controls. 5 ng of lambda DNA was added as quality control for cfDNA extraction.
(XLSX)

S1 Appendix. All repeat sequences identified using RepeatExplore on the Galaxy platform, based on *E. granulosus* and *E. multilocularis* sequencing data downloaded from NCBI.
(FASTA)

Acknowledgments

The authors wish to thank the study participants for their cooperation, the next-generation sequencing laboratory at the Center for Medical Genetics of Central South University for technical assistance, and Dr. Xuyu Cai for advise on cfDNA manipulation. We are particularly indebted to the reviewers for their critical suggestions and great help in improving the language of the manuscript.

Author Contributions

Conceptualization: Zhuohua Zhang.

Data curation: Zhengqing Wan, Xiaoqing Peng.

Formal analysis: Zhengqing Wan, Xiaoqing Peng, Lu Ma.

Funding acquisition: Zhuohua Zhang.

Investigation: Zhengqing Wan, Zhuohua Zhang.

Methodology: Zhengqing Wan, Zhuohua Zhang.

Project administration: Zhuohua Zhang.

Resources: Qingshan Tian, Shizheng Wu, Junqi Li, Jie Ling, Weigang Lv, Zhuohua Zhang.

Software: Zhengqing Wan, Xiaoqing Peng.

Supervision: Zhuohua Zhang.

Validation: Zhengqing Wan, Xiaoqing Peng.

Visualization: Zhengqing Wan.

Writing – original draft: Zhengqing Wan, Xiaoqing Peng, Zhuohua Zhang.

Writing – review & editing: Xiaoqing Peng, Binrong Ding, Jieqiong Tan, Zhuohua Zhang.

References

1. Deplazes P, Rinaldi L, Alvarez Rojas CA, Torgerson PR, Harandi MF, Romig T, et al. Global Distribution of Alveolar and Cystic Echinococcosis. *Adv Parasitol.* 2017; 95:315–493. <https://doi.org/10.1016/bs.apar.2016.11.001> PMID: 28131365
2. Echinococcosis Fact sheet. World Health Organization. March 2014.
3. McManus DP, Zhang W, Li J, Bartley PB. Echinococcosis. *Lancet.* 2003; 362(9392):1295–304. [https://doi.org/10.1016/S0140-6736\(03\)14573-4](https://doi.org/10.1016/S0140-6736(03)14573-4) PMID: 14575976
4. Holcman B, Heath DD. The early stages of Echinococcus granulosus development. *Acta Trop.* 1997; 64(1–2):5–17. [https://doi.org/10.1016/s0001-706x\(96\)00636-5](https://doi.org/10.1016/s0001-706x(96)00636-5) PMID: 9095285
5. Thompson RC. Biology and Systematics of Echinococcus. *Adv Parasitol.* 2017; 95:65–109. <https://doi.org/10.1016/bs.apar.2016.07.001> PMID: 28131366
6. Gharbi HA, Hassine W, Brauner MW, Dupuch K. Ultrasound examination of the hydatid liver. *Radiology.* 1981; 139(2):459–63. <https://doi.org/10.1148/radiology.139.2.7220891> PMID: 7220891
7. Polat P, Kantarci M, Alper F, Suma S, Koruyucu MB, Okur A. Hydatid disease from head to toe. *RadioGraphics.* 2003; 23(2):475–94; quiz 536–7. <https://doi.org/10.1148/rg.232025704> PMID: 12640161
8. Craig PS, McManus DP, Lightowler MW, Chabalgoity JA, Garcia HH, Gavidia CM, et al. Prevention and control of cystic echinococcosis. *Lancet Infect Dis.* 2007; 7(6):385–94. [https://doi.org/10.1016/S1473-3099\(07\)70134-2](https://doi.org/10.1016/S1473-3099(07)70134-2) PMID: 17521591
9. Zhang W, Wen H, Li J, Lin R, McManus DP. Immunology and immunodiagnosis of cystic echinococcosis: an update. *Clin Dev Immunol.* 2012; 2012:101895. <https://doi.org/10.1155/2012/101895> PMID: 22235225
10. de la Rue ML, Yamano K, Almeida CE, Iesbich MP, Fernandes CD, Goto A, et al. Serological reactivity of patients with Echinococcus infections (*E. granulosus*, *E. vogeli*, and *E. multilocularis*) against three antigen B subunits. *Parasitol Res.* 2010; 106(3):741–5. <https://doi.org/10.1007/s00436-009-1707-3> PMID: 20066435
11. Kalantari E, Bandehpour M, Pazoki R, Taghipoor-Lailabadi N, Khazan H, Mosaffa N, et al. Application of recombinant Echinococcus granulosus antigen B to ELISA kits for diagnosing hydatidosis. *Parasitol Res.* 2010; 106(4):847–51. <https://doi.org/10.1007/s00436-010-1726-0> PMID: 20143095
12. Hernandez-Gonzalez A, Muro A, Barrera I, Ramos G, Orduna A, Siles-Lucas M. Usefulness of four different Echinococcus granulosus recombinant antigens for serodiagnosis of unilocular hydatid disease (UHD) and postsurgical follow-up of patients treated for UHD. *Clin Vaccine Immunol.* 2008; 15(1):147–53. <https://doi.org/10.1128/CVI.00363-07> PMID: 17989342
13. Feng X, Wen H, Zhang Z, Chen X, Ma X, Zhang J, et al. Dot immunogold filtration assay (DIGFA) with multiple native antigens for rapid serodiagnosis of human cystic and alveolar echinococcosis. *Acta Trop.* 2010; 113(2):114–20. <https://doi.org/10.1016/j.actatropica.2009.10.003> PMID: 19836341
14. Vacirca D, Perdicchio M, Campisi E, Delunardo F, Ortona E, Margutti P, et al. Favourable prognostic value of antibodies anti-HSP20 in patients with cystic echinococcosis: a differential immunoproteomic

- approach. *Parasite Immunol.* 2011; 33(3):193–8. <https://doi.org/10.1111/j.1365-3024.2010.01264.x> PMID: 21306401
15. Zhang WB, Li J, Li Q, Yang D, Zhu B, You H, et al. Identification of a diagnostic antibody-binding region on the immunogenic protein EpC1 from *Echinococcus granulosus* and its application in population screening for cystic echinococcosis. *Clin Exp Immunol.* 2007; 149(1):80–6. <https://doi.org/10.1111/j.1365-2249.2007.03386.x> PMID: 17403055
 16. Marinova I, Nikolov G, Michova A, Kurdova R, Petrunov B. Quantitative assessment of serum-specific IgE in the diagnosis of human cystic echinococcosis. *Parasite Immunol.* 2011; 33(7):371–6. <https://doi.org/10.1111/j.1365-3024.2011.01292.x> PMID: 21480933
 17. Bauomi IR, El-Amir AM, Fahmy AM, Zalal RS, Diab TM. Evaluation of purified 27.5 kDa protoscolex antigen-based ELISA for the detection of circulating antigens and antibodies in sheep and human hydatidosis. *J Helminthol.* 2015; 89(5):577–83. <https://doi.org/10.1017/S0022149X14000479> PMID: 25006882
 18. Raether W, Hanel H. Epidemiology, clinical manifestations and diagnosis of zoonotic cestode infections: an update. *Parasitol Res.* 2003; 91(5):412–38. <https://doi.org/10.1007/s00436-003-0903-9> PMID: 13680371
 19. Gottstein B. An immunoassay for the detection of circulating antigens in human echinococcosis. *Am J Trop Med Hyg.* 1984; 33(6):1185–91. <https://doi.org/10.4269/ajtmh.1984.33.1185> PMID: 6391225
 20. Moosa RA, Abdel-Hafez SK. Serodiagnosis and seroepidemiology of human unilocular hydatidosis in Jordan. *Parasitol Res.* 1994; 80(8):664–71. <https://doi.org/10.1007/bf00932950> PMID: 7886035
 21. Li Y, Jia Z, Lv G, Wen H, Li P, Zhang H, et al. Detection of *Echinococcus granulosus* antigen by a quantum dot/porous silicon optical biosensor. *Biomed Opt Express.* 2017; 8(7):3458–69. <https://doi.org/10.1364/BOE.8.003458> PMID: 28717581
 22. Craig PS, Nelson GS. The detection of circulating antigen in human hydatid disease. *Ann Trop Med Parasitol.* 1984; 78(3):219–27. <https://doi.org/10.1080/00034983.1984.11811805> PMID: 6486929
 23. Sadjjadi SM, Sedaghat F, Hosseini SV, Sarkari B. Serum antigen and antibody detection in echinococcosis: application in serodiagnosis of human hydatidosis. *Korean J Parasitol.* 2009; 47(2):153–7. <https://doi.org/10.3347/kjp.2009.47.2.153> PMID: 19488422
 24. Devi CS, Parija SC. A new serum hydatid antigen detection test for diagnosis of cystic echinococcosis. *Am J Trop Med Hyg.* 2003; 69(5):525–8. PMID: 14695090
 25. Knapp J, Millon L, Mouzon L, Umhang G, Raoul F, Ali ZS, et al. Real time PCR to detect the environmental faecal contamination by *Echinococcus multilocularis* from red fox stools. *Vet Parasitol.* 2014; 201(1–2):40–7. <https://doi.org/10.1016/j.vetpar.2013.12.023> PMID: 24484767
 26. Al-Jawabreh A, Dumaidi K, Ereqat S, Nasereddin A, Al-Jawabreh H, Azmi K, et al. Incidence of *Echinococcus granulosus* in Domestic Dogs in Palestine as Revealed by Copro-PCR. *PLoS Negl Trop Dis.* 2015; 9(7):e0003934. <https://doi.org/10.1371/journal.pntd.0003934> PMID: 26181591
 27. Boubaker G, Macchiaroli N, Prada L, Cucher MA, Rosenzvit MC, Ziadinov I, et al. A multiplex PCR for the simultaneous detection and genotyping of the *Echinococcus granulosus* complex. *PLoS Negl Trop Dis.* 2013; 7(1):e2017. <https://doi.org/10.1371/journal.pntd.0002017> PMID: 23350011
 28. Wahlstrom H, Comin A, Isaksson M, Deplazes P. Detection of *Echinococcus multilocularis* by MC-PCR: evaluation of diagnostic sensitivity and specificity without gold standard. *Infect Ecol Epidemiol.* 2016; 6:30173. <https://doi.org/10.3402/iee.v6.30173> PMID: 26968153
 29. Knapp J, Umhang G, Pouille ML, Millon L. Development of a Real-Time PCR for a Sensitive One-Step Coprodiagnosis Allowing both the Identification of Carnivore Feces and the Detection of *Toxocara* spp. and *Echinococcus multilocularis*. *Appl Environ Microbiol.* 2016; 82(10):2950–8. <https://doi.org/10.1128/AEM.03467-15> PMID: 26969697
 30. Boufana B, Umhang G, Qiu J, Chen X, Lahmar S, Boue F, et al. Development of three PCR assays for the differentiation between *Echinococcus shiquicus*, *E. granulosus* (G1 genotype), and *E. multilocularis* DNA in the co-endemic region of Qinghai-Tibet plateau, China. *Am J Trop Med Hyg.* 2013; 88(4):795–802. <https://doi.org/10.4269/ajtmh.12-0331> PMID: 23438764
 31. Salant H, Abbasi I, Hamburger J. The development of a loop-mediated isothermal amplification method (LAMP) for *Echinococcus granulosus* [corrected] copro-detection. *Am J Trop Med Hyg.* 2012; 87(5):883–7. <https://doi.org/10.4269/ajtmh.2012.12-0184> PMID: 22987649
 32. Chaya D, Parija SC. Performance of polymerase chain reaction for the diagnosis of cystic echinococcosis using serum, urine, and cyst fluid samples. *Trop Parasitol.* 2014; 4(1):43–6. <https://doi.org/10.4103/2229-5070.129164> PMID: 24754027
 33. Jiang P, Lo YM. The Long and Short of Circulating Cell-Free DNA and the Ins and Outs of Molecular Diagnostics. *Trends Genet.* 2016; 32(6):360–71. <https://doi.org/10.1016/j.tig.2016.03.009> PMID: 27129983

34. Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J. Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell*. 2016; 164(1–2):57–68. <https://doi.org/10.1016/j.cell.2015.11.050> PMID: 26771485
35. Breitbach S, Tug S, Helmig S, Zahn D, Kubiak T, Michal M, et al. Direct quantification of cell-free, circulating DNA from unpurified plasma. *PLoS One*. 2014; 9(3):e87838. <https://doi.org/10.1371/journal.pone.0087838> PMID: 24595313
36. Breitbach S, Sterzing B, Magallanes C, Tug S, Simon P. Direct measurement of cell-free DNA from serially collected capillary plasma during incremental exercise. *J Appl Physiol* (1985). 2014; 117(2):119–30.
37. Volik S, Alcaide M, Morin RD, Collins C. Cell-free DNA (cfDNA): Clinical Significance and Utility in Cancer Shaped By Emerging Technologies. *Mol Cancer Res*. 2016; 14(10):898–908. <https://doi.org/10.1158/1541-7786.MCR-16-0044> PMID: 27422709
38. Kamat AA, Baldwin M, Urbauer D, Dang D, Han LY, Godwin A, et al. Plasma cell-free DNA in ovarian cancer: an independent prognostic biomarker. *Cancer*. 2010; 116(8):1918–25. <https://doi.org/10.1002/ncr.24997> PMID: 20166213
39. Lo YM, Zhang J, Leung TN, Lau TK, Chang AM, Hjelm NM. Rapid clearance of fetal DNA from maternal plasma. *Am J Hum Genet*. 1999; 64(1):218–24. <https://doi.org/10.1086/302205> PMID: 9915961
40. Crowley E, Di Nicolantonio F, Loupakis F, Bardelli A. Liquid biopsy: monitoring cancer-genetics in the blood. *Nat Rev Clin Oncol*. 2013; 10(8):472–84. <https://doi.org/10.1038/nrclinonc.2013.110> PMID: 23836314
41. Lo YM, Hjelm NM, Fidler C, Sargent IL, Murphy MF, Chamberlain PF, et al. Prenatal diagnosis of fetal RhD status by molecular analysis of maternal plasma. *N Engl J Med*. 1998; 339(24):1734–8. <https://doi.org/10.1056/NEJM199812103392402> PMID: 9845707
42. Chan KC, Jiang P, Sun K, Cheng YK, Tong YK, Cheng SH, et al. Second generation noninvasive fetal genome analysis reveals de novo mutations, single-base parental inheritance, and preferred DNA ends. *Proc Natl Acad Sci U S A*. 2016; 113(50):E8159–E68. <https://doi.org/10.1073/pnas.1615800113> PMID: 27799561
43. Lv W, Wei X, Guo R, Liu Q, Zheng Y, Chang J, et al. Noninvasive prenatal testing for Wilson disease by use of circulating single-molecule amplification and resequencing technology (cSMART). *Clin Chem*. 2015; 61(1):172–81. <https://doi.org/10.1373/clinchem.2014.229328> PMID: 25376582
44. Guo S, Diep D, Plongthongkum N, Fung HL, Zhang K, Zhang K. Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. *Nat Genet*. 2017; 49(4):635–42. <https://doi.org/10.1038/ng.3805> PMID: 28263317
45. Li Z, Guo X, Tang L, Peng L, Chen M, Luo X, et al. Methylation analysis of plasma cell-free DNA for breast cancer early detection using bisulfite next-generation sequencing. *Tumour Biol*. 2016; 37(10):13111–9. <https://doi.org/10.1007/s13277-016-5190-z> PMID: 27449045
46. Sun K, Jiang P, Chan KC, Wong J, Cheng YK, Liang RH, et al. Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments. *Proc Natl Acad Sci U S A*. 2015; 112(40):E5503–12. <https://doi.org/10.1073/pnas.1508736112> PMID: 26392541
47. De Vlaminc I, Martin L, Kertesz M, Patel K, Kowarsky M, Strehl C, et al. Noninvasive monitoring of infection and rejection after lung transplantation. *Proc Natl Acad Sci U S A*. 2015; 112(43):13336–41. <https://doi.org/10.1073/pnas.1517494112> PMID: 26460048
48. De Vlaminc I, Valentine HA, Snyder TM, Strehl C, Cohen G, Luikart H, et al. Circulating cell-free DNA enables noninvasive diagnosis of heart transplant rejection. *Sci Transl Med*. 2014; 6(241):241ra77. <https://doi.org/10.1126/scitranslmed.3007803> PMID: 24944192
49. Schutz E, Fischer A, Beck J, Harden M, Koch M, Wuensch T, et al. Graft-derived cell-free DNA, a noninvasive early rejection and graft damage marker in liver transplantation: A prospective, observational, multicenter cohort study. *PLoS Med*. 2017; 14(4):e1002286. <https://doi.org/10.1371/journal.pmed.1002286> PMID: 28441386
50. Snyder TM, Khush KK, Valentine HA, Quake SR. Universal noninvasive detection of solid organ transplant rejection. *Proc Natl Acad Sci U S A*. 2011; 108(15):6229–34. <https://doi.org/10.1073/pnas.1013924108> PMID: 21444804
51. Akirav EM, Lebastchi J, Galvan EM, Henegariu O, Akirav M, Ablamunits V, et al. Detection of beta cell death in diabetes using differentially methylated circulating DNA. *Proc Natl Acad Sci U S A*. 2011; 108(47):19018–23. <https://doi.org/10.1073/pnas.1111008108> PMID: 22074781
52. Lehmann-Werman R, Neiman D, Zemmour H, Moss J, Magenheimer J, Vaknin-Dembinsky A, et al. Identification of tissue-specific cell death using methylation patterns of circulating DNA. *Proc Natl Acad Sci U S A*. 2016; 113(13):E1826–34. <https://doi.org/10.1073/pnas.1519286113> PMID: 26976580

53. Olsen JA, Kenna LA, Tapon RC, Spelios MG, Stecker MM, Akirav EM. A Minimally-invasive Blood-derived Biomarker of Oligodendrocyte Cell-loss in Multiple Sclerosis. *EBioMedicine*. 2016; 10:227–35. <https://doi.org/10.1016/j.ebiom.2016.06.031> PMID: 27381476
54. Grumaz S, Stevens P, Grumaz C, Decker SO, Weigand MA, Hofer S, et al. Next-generation sequencing diagnostics of bacteremia in septic patients. *Genome Med*. 2016; 8(1):73. <https://doi.org/10.1186/s13073-016-0326-8> PMID: 27368373
55. Weerakoon KG, McManus DP. Cell-Free DNA as a Diagnostic Tool for Human Parasitic Infections. *Trends Parasitol*. 2016; 32(5):378–91. <https://doi.org/10.1016/j.pt.2016.01.006> PMID: 26847654
56. Dwivedi DJ, Toltl LJ, Swystun LL, Pogue J, Liaw KL, Weitz JI, et al. Prognostic utility and characterization of cell-free DNA in patients with severe sepsis. *Crit Care*. 2012; 16(4):R151. <https://doi.org/10.1186/cc11466> PMID: 22889177
57. Brunetti E, Kern P, Vuitton DA, Writing Panel for the W-I. Expert consensus for the diagnosis and treatment of cystic and alveolar echinococcosis in humans. *Acta Trop*. 2010; 114(1):1–16. <https://doi.org/10.1016/j.actatropica.2009.11.001> PMID: 19931502
58. Novak P, Neumann P, Pech J, Steinhaisl J, Macas J. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics*. 2013; 29(6):792–3. <https://doi.org/10.1093/bioinformatics/btt054> PMID: 23376349
59. Aziz A, Zhang W, Li J, Loukas A, McManus DP, Mulvenna J. Proteomic characterisation of *Echinococcus granulosus* hydatid cyst fluid from sheep, cattle and humans. *J Proteomics*. 2011; 74(9):1560–72. <https://doi.org/10.1016/j.jprot.2011.02.021> PMID: 21362505
60. Baraquin A, Hervouet E, Richou C, Flori P, Peixoto P, Azizi A, et al. Circulating cell-free DNA in patients with alveolar echinococcosis. *Mol Biochem Parasitol*. 2018; 222:14–20. <https://doi.org/10.1016/j.molbiopara.2018.04.004> PMID: 29679605