medRxiv preprint doi: https://doi.org/10.1101/2024.11.10.24317049; this version posted November 13, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in percetuity.

perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

1

2

Non-invasive Preimplantation Genetic Testing of Embryonic Genome in Spent Culture Medium

Lei Huang^{1,2#*}, Yangyun Zou^{3#}, Ruiqi Zhang^{4#}, Jin Huang^{5,6,7,8#}, Guangjun Yin^{1,2,9},
Quangui Wang⁹, Yingying Xia³, Jialin Jia^{5,6,7,8}, Zeyu Wu¹⁰, Dandan Cao⁹, Weiliang
Song^{1,2}, Yaqiong Tang^{1,2,9}, Kai Liu^{1,2,9}, Xiaoran Chai^{1,2}, Guo-Bo Chen¹¹, Sijia Lu^{3*}, Hao
Ge^{1,4*}, Jie Qiao^{1,2,5,6,7,8,12*} and Xiaoliang S. Xie^{1,2,9*}

- 7
- Biomedical Pioneering Innovation Center (BIOPIC), Peking University, Beijing,
 100871, China
- Beijing Advanced Innovation Center for Genomics, Peking University, Beijing,
 100871, China
- 12 3. Yikon Genomics Co., Ltd., Shanghai, 310120, China

4. Beijing International Center for Mathematical Research (BICMR), Peking
 University, Beijing, 100871, China

- State Key Laboratory of Female Fertility Promotion, Centre for Reproductive
 Medicine, Department of Obstetrics and Gynecology, Peking University Third
 Hospital, Beijing 100191, China
- 18 6. National Clinical Research Center for Obstetrics and Gynecology, Peking
 19 University Third Hospital, Beijing 100191, China
- Key Laboratory of Assisted Reproduction (Peking University), Ministry of
 Education, Beijing 100191, China
- 8. Beijing Key Laboratory of Reproductive Endocrinology and Assisted Reproductive
 Technology, Beijing 100191, China
- 24 9. Changping Laboratory, Beijing, China
- 25 10. School of Mathematical Sciences, Peking University, Beijing, 100871, China
- 11. Center for Reproductive Medicine, Department of Genetics and Genomic Medicine,
 and Clinical Research Institute, Zhejiang Provincial People's Hospital, People's
- 28 Hospital of Hangzhou Medical College, Hangzhou, Zhejiang, 310024, China
- 12. Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, 100871,
 China
- [#]These authors contribute equally to the work.
- 32 *Corresponding authors: X. S. Xie (sunneyxie@biopic.pku.edu.cn), J. Qiao
- 33 (jie.qiao@263.net), L. Huang(leihuangmed@gmail.com), H. Ge
- 34 (<u>haoge@pku.edu.cn</u>), S. Lu (lusijia@yikongenomics.com)
- 35

³⁶ NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

37

38

39

ABSTRACT

Traditionally, preimplantation genetic testing (PGT) for in vitro 40 fertilization (IVF) requires invasive trophectoderm (TE) biopsy, which 41 might be detrimental to the embryo. Recently proposed non-invasive PGT 42 (ni-PGT) utilizing cell-free DNA from spent embryo culture medium 43 (SCM) also faces serious challenges in accuracy, especially for monogenic 44 diseases (niPGT-M), due to trace DNA content, maternal cell 45 contamination, and high Allele Drop-Out (ADO) rates. In this study, an 46 improved linear single-cell whole genome amplification method and a 47 Bayesian linkage analysis model were used to enhance accuracy in niPGT-48 M. We achieved about 75% report rate across all samples and 100% 49 accuracy in the reported samples. Meanwhile, we reconstructed the 50 embryonic genome and calculated the risk of type II diabetes (T2D) via 51 niPGT-P, consistent well with those from TE biopsy samples. Our approach 52 alleviated the limitations of ni-PGT and offers a promising avenue for 53 advancing noninvasive PGT with potential clinical applications. 54

55

Keywords: non-invasive preimplantation genetic testing, monogenic
diseases, polygenic diseases, spent embryo culture medium, Bayesian
linkage analysis

59

60

INTRODUCTION

Preimplantation genetic testing (PGT) is widely used in clinical *in vitro* fertilization (IVF) to screen embryos for monogenic diseases, aneuploidy, or structural rearrangements. Next-generation sequencing (NGS)-based PGT allows simultaneous analysis of chromosome copy number, linkage, and targeted haplotyping for mutations [1-8]. However, traditional PGT requires trophectoderm (TE) biopsy, which poses risks to embryo health and is prone to sampling bias due to embryo mosaicism [9]. Recently,

minimally invasive or non-invasive methods using cell-free DNA (cfDNA) 68 from blastocoele fluid (BF) [10] or spent embryo culture medium (SCM) 69 [11,12] have been reported as alternatives for PGT with less risk and lower 70 price. Despite their potential, challenges like trace DNA amount, poor 71 DNA integrality, and complicated origin led to highly variable success rates 72 in the published references from different clinical centers [12-15]. 73 Although the success amplification rate in SCM was higher than that in BF 74 [13,16] using the same amplification method, it still does not meet the yield 75 required for clinical diagnosis application. 76

77

Even worse, genetic contents amplified from SCM can be potentially 78 affected by maternal cell contamination (MCC), as maternal DNA is often 79 longer than embryonic DNA [17,18] thus preferentially being amplified, 80 and by the presence of high levels of human serum albumin (HSA) protein 81 in culture medium. Current commercial single-cell amplification kits, such 82 as DOP-PCR, MDA, or MALBAC are not specially optimized to address 83 these challenges. In 2017, Linear Amplification via Transposon Insertion 84 (LIANTI) was developed [19], significantly reducing the false positive rate 85 of Single Nucleotide Polymorphisms (SNP) detection and much lower 86 amplification bias caused by fragment lengths. However, the yield of 87 LIANTI amplification for SCM is still insufficient. There is an urgent need 88 to develop an efficient and accurate amplification method to obtain 89 embryonic DNA information from SCM and/or BF. 90

91

Numerous reports have explored on non-invasive preimplantation genetic 92 testing for an euploidy (niPGT-A) [11,20,21], but studies on non-invasive 93 preimplantation genetic testing for monogenic diseases (niPGT-M) remain 94 limited [12,13,15,16, 22-24]. In niPGT-M studies using SCM, the genotype 95 concordance with trophectoderm biopsy results varied widely, from 21% 96 to 88%, limiting the clinical utility of niPGT-M [12-15]. The key criterion 97 for evaluating PGT-M is not only genotype concordance but also the 98 misdiagnosis rate, which is very low (<0.1%) with traditional methods 99 according to the ESHRE PGT consortium [25]. This emphasizes the need 100

101 to assess misdiagnosis risk in new niPGT-M approaches.

102

Accurate linkage analysis is crucial alongside the amplification method to 103 pinpoint disease-carrying chromosomes in the embryos and minimize 104 misdiagnosis, especially given the trace DNA content in SCM. Traditional 105 linkage analysis typically requires at least two fully informative markers 106 like STRs or SNPs, near the mutation site, with ADO rates below 5% [26]. 107 Ou et al. used 4 informative SNPs [15], and Liu et al. used 10 [12]. 108 However, since ADO rates in SCM or BF can exceed 5%, relying on a few 109 informative sites for linkage analysis may lead to misdiagnosis. 110 Informative SNP markers could be too far from the mutation site, 111 increasing recombination risk. Meanwhile, maternal cell contamination 112 can also raise the misdiagnosis rate. These challenges indicate that 113 traditional linkage analysis methods for niPGT-M may not be suitable for 114 clinical applications. Therefore, developing a highly accurate linkage 115 analysis method tailored for high ADO rates and maternal cell 116 contamination is crucial. 117

118

Furthermore, preimplantation genetic testing for polygenic disorders 119 (PGT-P) using polygenic risk scores (PRS) holds promise in reducing the 120 risk of polygenic diseases, such as coronary heart diseases, diabetes, and 121 cancers at the pre-conception stage [27-29]. Non-invasive PGT-P (niPGT-122 P) utilizing genetic information from SCM may, help alleviate some of the 123 ethical risk-benefit concerns by avoiding the harms of biopsy to the 124 embryos [30]. However, the challenge lies in the low coverage and inferior 125 genotype quality of embryo genomes due to limited genetic materials and 126 poor quality of cell-free DNA in SCM, which hinders the availability of 127 genome-wide genetic variants necessary for PRS evaluation. Therefore, 128 accurate whole genome reconstruction is crucial for the successful 129 implementation of niPGT-P. 130

131

In the present study, we improved the LIANTI method to amplify traceDNA from SCM, overcoming challenges posed by high-concentration

protein existence and amplification bias caused by DNA size difference. 134 We introduced a Bayesian-based linkage analysis method (BASE-niPGT-135 M) that incorporates high ADO rates and maternal cell contamination. This 136 method could accurately detect disease-carrying chromosomes and provide 137 a confidence level for each diagnosis. Furthermore, using the fragmented 138 and incomplete genomic data from SCM, we employed Pedigree-139 Population-based Imputation with Haploid Assumption (PPIHA) method 140 to reconstruct the embryonic genome. After reconstruction, the high 141 genome-wide coverage and accuracy of genetic information were obtained, 142 enabling PRS analysis. In summary, our study demonstrates high detection 143 accuracy for ni-PGT, offering a reliable and potentially applicable non-144 invasive method for preimplantation genetic testing intended for clinical 145 use. 146

147

148

RESULTS

Characterization of Cell-Free DNA in Spent Embryo Culture Medium
 Understanding the properties of DNA in SCM, such as fragment size

and DNA quantity, is crucial for selecting appropriate amplification 151 methods. Unlike biopsy samples, which typically contain one to three 152 complete cell nuclei, cfDNA in SCM or BF is present in low quantities and 153 of poor quality. Our results showed the cfDNA in the culture medium 154 exhibits highly variable DNA fragment size distribution and distinct DNA 155 ladder patterns (Extended data Fig. 1 and Supplementary Note S1), posing 156 a significant challenge for DNA amplification. Based on the above 157 evidences, an effective amplification method should minimize 158 amplification bias caused by DNA size variation. 159

Analysis of the sequencing results of cfDNA revealed low genome coverage in SCM, ranging from 10% to 50% (Extended data Fig. 2), in contrast to the genome coverage of biopsy samples, which typically exceeds 95%. This deficiency is not primarily attributed to the amplification-induced allelic dropout (ADO) but rather due to the largescale DNA information loss of paternal or maternal alleles, we called it

large fragment loss of haplotype (LFLoH), as shown in Extended data Fig.
3. This LFLoH phenomenon often results in the loss of informative SNPs
both upstream and downstream adjacent to the pathogenic site, which are
typically used in traditional linkage analysis. Informative SNPs are those
where the SNP is heterozygous in the parent carrying the disease and
homozygous in the other parent, and where the SNP remains heterozygous
in the embryo (Extended data Fig. 4).

Maternal cell contamination in embryo culture medium has long been 173 a challenge in diagnostic procedures, often leading to misdiagnosis. Using 174 informative SNPs with homozygous but different parental genotypes at the 175 whole-genome level, we observed MCC values ranging from -55.1% to 176 99.7% (Extended data Fig. 5). Negative values may result from copy 177 number gains in segments of paternal chromosomes or losses of maternal 178 chromosomal segments. Employing an iterative strategy within a 10 Mb 179 range upstream and downstream of pathogenic loci in our Bayesian linkage 180 analysis method (See below and Methods), we observed the modified 181 MCC values ranging from 0.1% to 99.7%. Aside from the absence of 182 negative values, with this region, 58 samples have an MCC of less than 2%, 183 indicating minimal maternal cell contamination, 43 samples have relatively 184 low levels of maternal cell contamination between 2% and 30%, and only 185 10 samples have maternal cell contamination levels between 30% and 80%, 186 with only 6 samples exceeding 80% (Extended data Fig. 5, Supplementary 187 Table S4). 188

189 It's important to note that the cfDNA in SCM may potentially 190 originate from cell apoptotic processes. Using a ligation-based method 191 (Supplementary Notes S1), we observed a clear DNA ladder in the culture 192 medium, as shown in Extended data Fig. 1.

193

194 Whole Genome Amplification Method for SCM

Given the unique characteristics of cfDNA in SCM, a whole genome amplification method with minimal fragment size bias is essential. We compared the fragment size amplification biases of the current WGA methods for PGT-M and found that LIANTI exhibited the least bias (short:

long = 1:5), while MDA and MALBAC showed a stronger preference for
amplifying longer fragments, with short: long amplification biases of 1:500
and 1:6000, respectively (Supplementary Table S1). LIANTI is a linear
amplification method known for its high SNP detection accuracy [19].

However, the original LIANTI method has low efficiency in amplifying such minute quantities of DNA, failing to meet clinical requirements. To address this, we improved the LIANTI method by modifying the lysis step, adding more transposons for insertion, introducing DNA primers during first-strand DNA synthesis, and implementing exponential amplification during second-strand synthesis, as depicted in Extended data Fig. 6 (see Methods for details).

The improved LIANTI method achieved a 100% amplification 210 success rate for clinical samples and demonstrated high SNP detection 211 accuracy, as shown in Extended data Fig. 7 and Supplementary Table S2. 212 We compared the sequencing results of cfDNA from BJ cell line culture 213 medium collected at different time points using the improved LIANTI 214 method and other scWGA methods. The improved LIANTI method 215 demonstrated superior amplification efficiency for cfDNA compared to the 216 other scWGA methods. 217

218

219 Linkage Analysis Method for Noninvasive PGT-M of Embryos

Figure 1 illustrates the complete workflow employed in this study. 220 Briefly, data collection was divided into two stages (Figure 1A): the Pre-221 phasing stage and the SCM genetics information acquisition stage. In the 222 Pre-phasing stage, family trio sequencing was performed using DNA from 223 the father, mother, and proband or alternatively from grandparents or 224 discarded embryos. During the SCM genetics information acquisition stage, 225 the DNA content in SCM was amplified using the improved LIANTI 226 method, and then was followed by sequencing and SNP calling (See Fig. 227 1A and Methods). 228

With these data in hand, we present a computational method BASEniPGT-M (BAyesian linkage analySis mEthod for non-invasive Preimplantation Genetic Testing of Monogenic disorders) (Fig. 1B,

Extended data Fig. 8 and Supplementary Note S2). This approach 232 addresses the challenges posed by MCC and LFLoH in SCM samples. The 233 Bayesian model takes the sequencing data of SCM and the phased 234 haplotypes of the parents, as input. Samples with extremely low quality are 235 filtered out. Unlike previous approaches for high-quality sequencing data, 236 which often predetermined the number of SNPs to be considered, our 237 model incrementally incorporates SNPs, starting from the disease-causing 238 mutation site. This process allows us to calculate the log-likelihood ratio 239 of inheriting disease-carrying chromosomes versus disease-free 240 chromosomes for each SNP subset, ultimately resulting in a log-likelihood 241 ratio curve (Fig. 1B, Extended data Fig. 8 and Supplementary Note S2). 242 This curve, with its distinct characteristics, enables us to classify the results 243 into four categories: Highly Confident, Moderately Confident, Possibly 244 confident, and Undetermined. In the first three categories, we can pinpoint 245 the inherited chromosome. 246

- 247
- 248

Case Study of our niPGT-M Method

We selected one case (Family 2) to illustrate the SNP distribution of 249 cfDNA and the results of our niPGT-M method, which consists of WGA 250 and NGS followed by BASE-niPGT-M. This case involved an autosomal 251 dominant genetic disease, where the husband carried a genetic variant for 252 Marfan syndrome, a heritable connective tissue disorder. This disorder is 253 known for its characteristic features such as elongated limbs, aortic 254 aneurysms, and mitral valve prolapse. Genetic diagnosis of the husband 255 revealed a mutation c.643C>T at the FBN1 gene, known to cause Marfan 256 syndrome. Eight embryos of the couple developed to the blastocyst stage, 257 and several trophectoderm cells were biopsied from each embryo for 258 traditional PGT-M. All eight SCM samples were collected, and cfDNA was 259 amplified using the improved LIANTI method. The linkage analysis results 260 of these eight SCMs are shown in Fig. 2A. The detected SNP distribution 261 of 2T-SCM-01 and 2T-SCM-02 appeared relatively scattered, with longer 262 intervals between informative SNPs (Fig. 2A), and the SNPs flanking the 263 causal mutation site indicated different haplotypes from the father (Fig. 2B), 264

leading to their classification as 'undetermined'. In contrast, 2T-SCM-03, 265 2T-SCM-04, 2T-SCM-05, 2T-SCM-06 and 2T-SCM-08 showed denser 266 SNP distributions, and SNPs upstream and downstream of the causal 267 mutation site indicated the same haplotype (Fig. 2B), resulting in a 268 classification of 'Highly confident'. The detected SNPs of SCM-07 were 269 dense, and the haplotype upstream and downstream was concordant. 270 However, an upstream SNP located relatively close to the causal mutation 271 site pointed to the opposite haplotype (Fig. 2B), leading to a classification 272 of 'moderately confident' due to our conservative diagnosis strategy. All 273 these results were consistent with those of trophectoderm biopsies, 274 demonstrating no misclassifications. 275

276

277 Validation Using all Matched Biopsy/ Whole Embryo.

To further evaluate the performance of our niPGT-M method, we compared the results obtained from all SCM samples to the gold standard, which was either TE-biopsy or discarded embryo PGT-M results (Table 1). We categorized the cases into three groups: autosomal dominant disease, autosomal recessive disease, and X-linked recessive disease. For autosomal recessive diseases, we compared each allele of embryos.

In the seven cases with autosomal dominant diseases, there were 45 284 SCM samples. For embryos that the father carried the causal mutation, 23 285 out of 28 SCMs (82.1% report rate) were successfully diagnosed. For 286 embryos that the mother carried the causal mutation, 11out of 17 SCMs 287 (64.7% report rate) were successfully diagnosed. In the five cases with 288 autosomal recessive disorders, 36 SCMs were analyzed. We successfully 289 detected the paternal allele in 25 SCMs (69.4% report rate) and the 290 maternal allele in 27 SCMs (75% report rate). For the two cases with X-291 linked recessive disorders, 8 out of 9 SCMs (88.9% report rate) were 292 diagnosed. Overall, we successfully detected 94 out of 126 alleles (74.6% 293 average report rate). The average report rate for paternal alleles was 75%, 294 while the average report rate for maternal alleles was 74.2%. 295

Importantly, all detected results were concordant with the gold standard, demonstrating a 100% accuracy rate of BASE-niPGT-M (Table

298 1).

299

300 Whole Genome Reconstruction of Embryos

The directly detected genetic information of embryos obtained from the SCM samples is much limited due to low amount of cfDNA materials released from blastocyst cells. We therefore developed a Pedigree-Population-based Imputation with Haploid Assumption (PPIHA) to acquire the missing genetic information. This approach allows us to reconstruct the whole genome information of embryos from SCM samples.

The initial density of detected sites in the raw data was 222±127 SNPs per megabase (Mb). However, after performing PPIHA (shown in Fig. 3A), the median coverage of SNPs significantly improved from 6.4% to 93.5%, and the density of imputed sites increased to 1479±554 SNPs/Mb (Fig. 3B and 3D).

To assess the accuracy of the predicted embryo's haplotype obtained 312 from SCM samples, we compared it with the haplotype of corresponding 313 discarded embryos or TE biopsy samples. When the MCC of SCM was low, 314 we observed a high similarity of haplotype configuration between SCM 315 and the corresponding TE biopsy. Conversely, SCM samples with high 316 MCC exhibited reduced haplotype similarity, especially in the maternal 317 haplotypes (Fig. 3C). For a more comprehensive exploration, we obtained 318 43 paired samples with whole genome sequencing data from both SCMs 319 and their corresponding discarded embryos or TE biopsy samples for 320 comparison. To avoid bias we removed paired samples with failed quality 321 control or aneuploid embryos for further analysis (Supplementary Table 322 S3). Paternal or Maternal haplotype concordance at each physical site 323 between SCM and discarded embryos/TE biopsy samples was applied to 324 measure the haplotype accuracy constructed in SCM (Fig. 3E). We found 325 that the concordance of maternal and paternal haplotypes in SCM versus 326 discarded embryos/TE was around 98.6% and 99.8% respectively, for 327 SCM samples without significant maternal cell contamination (Fig. 3E). 328 High levels of maternal cell contamination could largely impact the 329 haplotype construction of SCM, especially for maternal haplotypes (Fig. 330

331 **3**E).

Furthermore, after filtering out SNPs with Mendelian errors, we 332 compared the constructed genome-wide genotypes from SCM samples 333 with those of corresponding discarded embryos/TE biopsy samples. The 334 concordance of genotypes ranged from 79.2% to 99.9%, partially impacted 335 by maternal cell contamination in SCM (Fig. 3E). For those SCM without 336 significant maternal cell contamination, the average genotype accuracy 337 was approximately 96.9%. Additionally, we compared the constructed 338 genotypes from two SCM samples with the genotypes of corresponding 339 amniotic fluid samples, which exhibited concordance rates of 94.5% and 340 94.1%, respectively. 341

342

343 **Polygenic Risk Evaluation of Type II Diabetes for Embryos**

We used type II diabetes (T2D) as a complex disease model to assess 344 the feasibility of polygenic risk evaluation based on constructed whole 345 genome information of SCM. After applying various filters, including 346 removing SCM samples with high maternal cell contamination, 347 chromosomal abnormality or poor-quality, 25 SCM samples were left for 348 polygenic risk score (PRS) calculation of T2D. We employed a PRS model 349 of T2D from a previous study [31]. The PRS of T2D model involves a total 350 of 114 loci, where only 28.7% of the loci can be directly detected, 52.5% 351 being obtained through family-based genome reconstruction, and the 352 remaining 18.8% being imputed by population-based reconstruction (Fig. 353 3B, Fig. 4A). 354

When comparing the genotypes of SNPs in the T2D PRS model between the SCM samples and corresponding discarded embryos or TE biopsy samples, we found that median 97.4% of the SNPs were identical. Among these identical SNPs, 52.5% are homozygous REF genotypes, 24.9% are homozygous ALT genotypes, and 22.6% are heterozygous genotypes (Fig. 4B).

The PRS of T2D calculated from SCM showed a high level of concordance with those calculated from the corresponding discarded embryos, TE biopsy samples, or amniotic fluid, with an R-squared value

of 0.79 (Fig. 4C). In clinical practice, PRS percentiles are typically used to 364 evaluate the relative genetic risk of common diseases instead of raw PRS 365 scores. To determine an individual's PRS percentile for a particular disease, 366 a reference population comprising both patients (e.g., with T2D) and 367 healthy individuals is needed to establish the PRS distribution, which then 368 allows each individual to be placed within a specific percentile of the 369 distribution. For example, if an individual's PRS is in the 90th percentile, 370 it means he/she has a higher PRS than 90% of the reference population. 371 Generally, the higher the PRS percentile, the greater the risk of a genetic 372 disease. We used a reference population from the UK Biobank 373 (https://www.ukbiobank.ac.uk/) to construct the PRS distribution for T2D 374 and calculated each embryo's PRS percentile. We observed a correlation of 375 T2D PRS percentiles between SCM samples and the corresponding 376 discarded embryos, TE biopsy samples, or amniotic fluid, indicated by an 377 R-squared value of 0.68 (Fig. 4D). 378

379

380

DISCUSSION

introduced a method for In this study, we non-invasive 381 preimplantation genetic testing, addressing key challenges in the accuracy 382 of genetic assessments using SCM. We improved a linear single-cell whole 383 genome amplification method to successfully amplify DNA content in 384 SCM and introduced a Bayesian linkage analysis model to improve the 385 accuracy of ni-PGT for monogenic diseases, matching the reliability of 386 invasive procedures. Our approach advances the potential clinical 387 implementation of ni-PGT, reducing the risks associated with traditional 388 PGT while maintaining high accuracy. 389

PGT-P holds promise for reducing the incidence of complex genetic diseases at the population level. However, its widespread use is hindered by controversies regarding ethical concerns, scientific validity, and practical issues, particularly the potential harm of the biopsy procedure to embryos if the benefits of PGT-P are not conclusive. Non-invasive PGT-P, which detects genetic materials from embryos through SCM, may

accelerate scientific research and clinical implementation by significantly 396 reducing potential risks to embryos. We achieved comprehensive whole-397 genome construction of embryos, accurately assessing the risk of polygenic 398 diseases (niPGT-P) in line with results from trophectoderm (TE) biopsy 399 samples. Our PPIHA approach to reconstruct whole genome information 400 of embryos from SCM samples fully considers the genetic characteristics 401 of SCM specimens, such as LFLoH, which violates the general biological 402 principles of diploidy, and the low amount and quality of cfDNA released 403 from blastocysts. The high genome-wide coverage and accuracy of genetic 404 information in SCM samples constructed through PPIHA enable non-405 invasive assessment of polygenic disease risk in embryos, advancing the 406 prevention and control of common diseases to the pre-implantation stage. 407

Maternal cell contamination from cumulus cells or polar bodies in SCM is an issue that requires attention, which would impact the accuracy of niPGT-M and niPGT-P. Special techniques for the removal of maternal cell contamination are necessary to be developed in the future. At current phase, we created a quantitative estimation of MCC rate which plays a warning role when it is high, and ensures accurate ni-PGT results when it is not high.

In contrast to TE samples that contain alleles from two haplotypes, 415 SCM only provides alleles from a single haplotype in many genomic 416 regions, resulting in a high rate of allelic dropout and random segments of 417 single haplotypes in the SCM data. To address this issue, we developed a 418 specialized approach to first estimate the haplotype status of each SNP, 419 specifically whether only the parental or maternal haplotype detected. 420 Assuming the independence of haplotype status among measured SNPs 421 would introduce an excessive number of parameters, leading to the risk of 422 overfitting. Therefore we have adopted a cautious approach: the default 423 assumption for the haplotype status of each SNP is set to be "both parental 424 haplotypes are detected", unless the SNP data strongly contradicts this 425 default. 426

The accuracy of haplotype pre-phasing of the parents is important for both niPGT-M and niPGT-P. In cases where no proband is available,

429 sequencing data from discarded embryos must be used for pre-phasing. 430 However, the data quality, particularly the coverage, obtained from 431 discarded embryos is often limited, which can hinder the accuracy of 432 haplotype phasing results for the parents. Improved precision in haplotype 433 phasing of the parents, possibly achievable through more advanced 434 sequencing technologies, would enhance the accuracy of niPGT-M and 435 niPGT-P.

In the context of clinical applications, our methodology extends 436 beyond merely estimating inherited chromosomes in the embryo; it also 437 provides a measure of confidence in this estimation. By utilizing a 438 Bayesian model, the log-likelihood ratio curve enables a nuanced 439 classification of our diagnoses. This approach is crucial for aiding both 440 medical professionals and parents in determining the most suitable course 441 of action for the embryo. The precision of our method, combined with the 442 refined classification outputs, significantly enhances the reliability of 443 identifying healthy embryos. This enhancement underscores the potential 444 translational impact of our work, promising increased report rates in 445 pregnancies and the delivery of healthy infants. 446

- 447
- 448

REFERENCES

- Yan, L. et al. Live births after simultaneous avoidance of monogenic diseases and chromosome abnormality by next-generation sequencing with linkage analyses.
 Proc Natl Acad Sci U S A . 112(52):15964-9 (2015)
- 452 2. Esteki, Z.M. et al. Concurrent whole-genome haplotyping and copy-number
 453 profiling of single cells. *Am J Hum Genet*. 96(6):894-912 (2015)
- 3. Xiong, L. et al. Bayesian model for accurate MARSALA (mutated allele revealed
 by sequencing with aneuploidy and linkage analyses). *JAssist Reprod Genet.* 36(6):
 1263–1271 (2019)
- 4. Backenroth, D. et al. Haploseek: a 24-hour all-in-one method for preimplantation
 genetic diagnosis (PGD) of monogenic disease and aneuploidy. *Genet Med.*21(6):1390-1399 (2019)
- Masset, H. et al. Single-cell genome-wide concurrent haplotyping and copy-number
 profiling through genotyping-by-sequencing. *Nucleic Acids Res.* 50(11):e63 (2022)
- 462 6. Yan, L. et al. Chinese experts' consensus guideline on preimplantation genetic
 463 testing of monogenic disorders. *Hum Reprod.* 38(Supplement_2):ii3-ii13 (2023)

- 7. Backenroth, D. et al. SHaploseek is a sequencing-only, high-resolution method for 464 comprehensive preimplantation genetic testing. Sci Rep. 13(1):18036 (2023) 465 8. De Coster, T. et al. Genome-wide equine preimplantation genetic testing enabled by 466 simultaneous haplotyping and copy number detection. Sci Rep. 14(1):2003 (2024) 467
- 9. McCoy, R.C. Mosaicism in preimplantation human embryos: when chromosomal 468 abnormalities are the norm. Trends Genet. 33(7):448-463 (2017) . 469
- 10. Palini, S. et al. Genomic DNA in human blastocoele fluid. Reprod Biomed Online 470 26, 603-610 (2013). 471
- 11. Xu, J. et al. Noninvasive chromosome screening of human embryos by genome 472 sequencing of embryo culture medium for in vitro fertilization. Proc Natl Acad Sci 473 113, 11907–11912 (2016). 474
- 12. Liu, W.O. et al. Non-invasive preimplantation aneuploidy screening and diagnosis 475 of beta thalassemia IVSII654 mutation using spent embryo culture medium. Ann 476 *Med* 49, 319–328 (2017) 477
- 13. Capalbo, A. et al. Diagnostic efficacy of blastocoel fluid and spent media as sources 478 of DNA for preimplantation genetic testing in standard clinical conditions. Fertil 479 Steril 110, 870-879 (2018). 480
- 481 14. Cimadomo, D. et al. The dawn of the future: 30 years from the first biopsy of a human embryo. The detailed history of an ongoing revolution. Human 482 Reproduction Update. 26, 453-73 (2020). 483
- 15. Ou, Z. et al. Improved non-invasive preimplantation genetic testing for beta-484 thalassemia using spent embryo culture medium containing blastocoelic fluid. 485 Frontiers in endocrinology 12, 1941 (2022). 486
- 16. Galluzzi, L. et al. Extracellular embryo genomic DNA and its potential for 487 genotyping applications. Futur Sci OA 1, FSO62 (2015). 488
- 17. Lo, Y.M.D. et al. Maternal plasma DNA sequencing reveals the genome-wide 489 genetic and mutational profile of the fetus. Sci. Transl. Med. 2, 61ra91 (2010) 490
- 18. Yu, S.C.Y. et al. Size-based molecular diagnostics using plasma DNA for 491 noninvasive prenatal testing. Proc. Natl. Acad. Sci. USA 111, 8583-8588 (2014) 492
- 19. Chen, C. et al. Single-cell whole-genome analyses by Linear Amplification via 493 Transposon Insertion (LIANTI). Science 356, 189-94 (2017). 494
- 495 20. Huang, L. et al. Noninvasive preimplantation genetic testing for an uploidy in spent medium may be more reliable than trophectoderm biopsy. Proc. Natl. Acad. Sci. 496 USA 116, 14105-14112 (2019) 497
- 21. Leaver, M. & Wells, D. Non-invasive preimplantation genetic testing (niPGT): the 498 next revolution in reproductive genetics? Human reproduction update 26, 16-42 499 (2020).500
- 22. Wu, H. et al. Medium-based noninvasive preimplantation genetic diagnosis for 501

502	human α-thalassemias-SEA. Medicine (Baltimore) 94, e669 (2015).
503	23. Zhang, Y. et al. Molecular analysis of DNA in blastocoele fluid using next-
504	generation sequencing. J Assist Reprod Genet 33, 637-645 (2016).
505	24. Shangguan, T. et al. Detection and analysis of DNAmaterial in human blastocoel
506	fluid. Biomed Genet Genomics 2, 1-5 (2017).
507	25. De Rycke, M. et al. ESHRE PGD Consortium data collection XIV-XV: Cycles from
508	January 2011 to December 2012 with pregnancy follow-up to October 2013. Hum.
509	<i>Reprod.</i> 32, 1974–1994 (2017).
510	26. Carvalho, F. et al. ESHRE PGT Consortium good practice recommendations for the
511	detection of monogenic disorders. Human Reproduction Open 2020(3), hoaa018
512	(2020).
513	27. Treff, N.R. et al. Utility and first clinical application of screening embryos for
514	polygenic disease risk reduction. Frontiers in Endocrinology 10, 845 (2019).
515	28. Kumar, A. et al. Whole-genome risk prediction of common diseases in human
516	preimplantation embryos. <i>Nature Medicine</i> 28, 513–516 (2022).
517	29. Lennon, N.J. et al. Selection, optimization and validation of ten chronic disease
518	polygenic risk scores for clinical implementation in diverse US populations. <i>Nature</i>
519	Medicine 30, 480–487 (2024).
520	50. Turley, P. et al. Problems with Using Polygenic Scores to Select Embryos. New
521	England Journal of Medicine 585, 78–80 (2021).
522	in Europeans, Diabatas 66, 2888, 2002 (2017)
525 521	in Europeans. <i>Diubeles</i> 00, 2888–2902 (2017).
524	
525	METHODS
526	Institutional Review Board Approval.
527	This study was approved by the Peking University Third Hospital
528	Ethical Review Committee (Approval no. 2020SZ-005). The
529	patients/participants provided written informed consent to participate in
530	this study.
531	Study Subjects.
532	Patients (n=14 couples) using ICSI and PGT-M with TE biopsy for
533	preventing their inherited genetic diseases at Peking University Third
534	Hospital during the period from September 2019 to April 2021 were

included. All of the 14 patients had genetic testing reports.

536 **ART Laboratory Protocols and Blastocyst Biopsy for PGT-M.**

537 Standard protocols were used for ICSI. Embryos were cultured

individually in 25-µL microdrops of the equilibrated Vitrolife G-series 538 media with human serum albumin (HSA) (LifeGlobal) overlain with 539 mineral oil in incubators with 6% CO2 and 5% O2 at 37 °C. Embryos were 540 evaluated on day 5 or 6 for trophectoderm (TE) biopsy. A few TE cells 541 from each hatching blastocyst were biopsied and sent to PGT lab for PGT-542 M testing. 543

Collection of Culture Medium Sample and Discarded Embryo 544

Immediately after the blastocysts were transferred to another dish for 545 biopsy, 20-µL spent embryo culture media were removed from each of the 546 residual spent medium drops. The discarded embryos and their 547 corresponding spent embryo culture media samples were also collected for 548 analysis. Each discarded embryo was gently moved with a pipette tip to the 549 edge of its microdrop and then removed and transferred into a RNase-550 DNase-free PCR tube containing 3 µL of lysis buffer. Pipette tips were 551 changed between collections of each sample to avoid cross-sample 552 contamination. Media drops incubated and collected under identical 553 conditions to those used for blastocyst culture, but never containing 554 embryos, served as negative controls. All samples were immediately frozen 555 after collection and stored at -80 °C until analyzed. 556

Spent Embryo Culture Medium Lysis and Discarded Embryo Lysis 557

The frozen 20-µL spent embryo culture medium samples were thawed 558 and gently mixed, and 10.8 µL was removed to a PCR tube (Maximum 559 Recovery, Axygen) containing 1.2 µL 10x lysis buffer (1x lysis buffer: 60 560 mM Tris-Ac pH 8.3, 2 mM EDTA pH 8.0, 15 mM DTT, 0.5 µM carrier 561 ssDNA (5'-TCAGGTTTTCCTGAA-3', Thermo Fisher Scientific oligo 562 with PAGE purification)), spun down. It was heated at 75°C for 30 min, 563 then incubated at 75°C for 30 min. 0.5 µL 35mg/mL QIAGEN protease 564 (dissolved in water and stored at 4°C) was added, spun down, and 565 incubated at 55°C for 2 hrs followed by 80°C for 30 min. The resulting 566 lysate in PCR tubes could be immediately subject to improved linear 567 transposon-based amplification as described herein or stored in a freezer 568 for later use. 569

570

The frozen discarded embryos were thawed and heated at 75°C for 30

571 min. Then the discarded embryo samples were lysed by adding 0.5 μ L

572 5mg/mL Qiagen Protease and heating (2 h at 55 °C and 30 min at 80 °C)

573 in 3 μ L of lysis buffer.

574 Whole Genome Amplification

575 Transposon

DNA

(5'/Phos/CTGTCTCTTATACACATCTGAACAGAATTTAATACGACTC 576 ACTATAGGGAGATGTGTATAAGAGACAG-3', Thermo Fisher 577 Scientific oligo with PAGE purification) was annealed by gradual cooling 578 in annealing buffer (20 mM Tris-Ac pH 8.3, 50 mM NaCl, 2 mM EDTA 579 pH 8.0) into a 1.5 µM self-looping structure. The 1.5 µM annealed 580 transposon DNA was then mixed with an equal volume of $\sim 1 \mu M$ Tn5 581 transposase (Lucigen, EZ-Tn5[™] Transposase) and incubated at room 582 temperature for 30 min, dimerizing into transposomes with a final 583 concentration of ~0.25 μ M. The transposomes can be stored at -20°C for a 584 long time, and each single-cell transposonbased amplification needs ~ 0.5 585 μL transposome. 586

Starting with the culture medium lysate, a 20 µL transposition mixture 587 was assembled in the buffer containing 2.5 mM MgCl2 and 18.75 nM 588 transposome. The transposition reaction was carried out at 55°C for 12 min. 589 Then 0.84 µL mixture of EDTA and ssDNA was added to each tube and 590 was incubated at 68 °C for 30 min. After removing transposase reaction, 591 0.2 µL Q5 HF DNA Polymerase (New England Biolabs) was added and 592 was heated at 73°C for 45 seconds in the presence of 2 mM MgCl2 and 593 200 µM dNTPs for fragments end filling and extension. 0.5 µL 4mg/mL 594 Qiagen protease was added and the PCR tube was heated at 50 °C for 1 595 hour in the presence of 4 mM EDTA to inactivate DNA polymerase, 596 followed by protease heat inactivation by 77°C for 20 min in the presence 597 of 450 mM NaCl in a total volume of ~25 µL. DNA fragments were 598 amplified to RNAs in a 90 µL T7 in vitro transcription reaction mixture 599 overnight at 37 °C as described in Cheng et al's paper [19]. 600

601 The next day, 10 μ L 0.5M EDTA was added to each tube and RNAs 602 were column purified (Zymo Research). 18 μ L RNAs were transferred to 603 a PCR tube containing 3.1 μ L mixture of 6.5mM each dNTPs, 9.7 μ M

carrier ssDNA (5'-TCAGGTTTTCCTGAA-3', Thermo Fisher Scientific 604 oligo with PAGE purification) and 3.2 U/µL SUPERase In RNase Inhibitor 605 (Invitrogen). The denaturation incubation was at 70 °C 1min, 90 °C 15s, 606 and followed by ice quenching. The reverse transcription reaction for the 607 first stand was carried out in a 30 µL SuperScript IV reserve reaction 608 system including 0.67 mM each dNTPs, 0.6 U/µL SUPERase In RNase 609 Inhibitor (Invitrogen) and 6 U/µL SuperScript IV Reserve Transcriptase 610 SuperScript IV buffer, with primer (Invitrogen) in the 5'-611 AGATGTGTATAAGAGACAG-3', Thermo Fisher Scientific oligo with 612 PAGE purification, and the incubation program was 25°C 1 min, 37°C 1 613 min, 42°C 1 min, 50°C 1 min, 55°C 15 min, 60°C 10 min, 65°C 12 min, 614 70°C 8 min, 75°C 5 min, 80°C 10 min. RNA was then removed by 615 incubation at 37°C for 30 min with 10 ng/µL affinity-purified RNase A 616 (Invitrogen) and 0.08 U/µL RNase H (New England Biolabs). Second 617 strand synthesis was carried out in a 100 µL Q5 DNA polymerase system 618 (New England Biolabs, 1X Q5 reaction buffer, 1X Q5 High GC enhancer, 619 200 µM dNTPs, 0.5 µM primer, 0.02 U/µL Q5 DNA polymerase), with the 620 primer 5'-NNNNNNNGGGAGATGTGTATAAGAGACAG-3', Thermo 621 Fisher Scientific oligo with PAGE purification. Each tube was heated at 622 98°C for 30 s, then using 10 cycles of 10 s at 98°C, 30s at 58°C, 30s at 623 60°C, 30s at 65°C, 2.5 min at 70°C, then 72°C 6 min for strand extension. 624 The resulted amplicons were column purified into 23 µL elution buffer and 625 stored at -20°C. The DNA concentration of the product after amplification 626 was measured using a Qubit 2.0 fluorometer (ThermoFisher Scientific) 627 with the Qubit dsDNA HS Assay kit (Life Technologies). 628

629 Starting with the discarded embryo lysate, DNA was amplified 630 following the LIANTI amplification method described in [19], except the 631 DNA primer 19_1 5'-AGATGTGTATAAGAGACAG-3', Thermo Fisher 632 Scientific oligo with PAGE purification was added in the reaction of the 633 first strand cDNA synthesis.

634 Sequencing and Data Analysis.

635 Upon library preparation, the amplicons were subject to a 636 conventional sonication process (Covaris S2) to the length required by

sequencing platforms. The final insert size was suitable for 2x150 bp pair-637 end Illumina sequencing. Library preparation was performed by NEBNext 638 Ultra II DNA Library Prep Kit for Illumina (New England Biolabs), 639 following the instructions of the manufacturer and skipping the optional 640 size selection step. For the gDNA of blood samples, all the samples were 641 subject to a conventional sonication process (Covaris S2). Library 642 preparation was performed by a PCR-free Library Prep Kit, NEBNext 643 Ultra II DNA Library Prep Kit for Illumina (NEB #E7645S/L). 644

The bulk samples, culture medium samples, and discarded embryo samples were sequenced on the Illumina NovaSeq 6000 System or Illumina HiSeq HiSeq 4000 platform with 2x150 bp pair-end sequencing. Sequencing lanes were shared by 24 samples with NEBNext Indexes. Each gDNA sample was sequenced to generate ~90 Gb raw data and each culture medium sample or discarded embryo was sequenced to generate ~30 Gb raw data.

652 Mapping and SNP Calling

Raw paired-end reads were trimmed adapters and low-quality ends by 653 cutadapt. Clean reads were mapped to human reference genome hg19 with 654 BWA mem. Each BAM file was marked duplicates and sorted by 655 MarkDuplicatesSpark of GATK 4.2.0. Base quality score recalibration 656 (BQSR) was performed using BaseRecalibrator to generate a recalibration 657 table, followed by ApplyBQSR to adjust the base quality accordingly. For 658 variant calling, GATK HaplotypeCaller produced a genomic variant call 659 format (GVCF) file for each sample. GenotypeGVCFs was utilized to 660 perform joint genotyping across all samples. SNPs were filtered using 661 variant quality score recalibration (VQSR) with HapMap 3.3, Omni 2.5, 662 1000 Genome phase I, and dbSNP 151 as SNP training sets. A 99% 663 sensitivity threshold was chosen to filter SNPs accurately. 664

665 We exclusively retain biallelic SNPs for subsequent analysis. To 666 deduce the two parental haplotypes, we employ Plink 1.9 with the option 667 "--mendel" to identify sites with Mendelian errors.

668

669 Parental haplotype phasing

For parental haplotype phasing, genetic information of pedigrees such as 670 sibling embryos or other family members are needed. Likelihood-based 671 haplotyping approach using Hidden Markov Model (HMM) and dynamic 672 programming Viterbi algorithm were used to determine the most likely 673 haplotype configuration of parents [32,33]. The transition probability of 674 haplotype state changes between consecutive loci in the HMM was 675 calculated from the recombination fraction obtained from the genetic 676 distance map of the 1000 Genomes Project Phase 3 references 677 (https://www.internationalgenome.org/). To address the ADO issue 678 stemming from scWGA bias during parental haplotyping with sibling 679 embryos, we filtered out sites that violated Mendelian rules or phased 680 variants that did not adhere to chromosome interference theory [34]. Each 681 chromosome of parent was then independently processed until the parental 682 haplotypes of all chromosomes were phased. 683

- 684
- 685

BASE-niPGT-M: a BAyesian linkage analySis mEthod for niGPT-M

686

Bayesian model. We present BASE-niPGT-M, a Bayesian linkage analysis 687 method tailored for niPGT-M. Our approach incorporates the MCC rate 688 and haplotype status of each SNP into the likelihood function. To enhance 689 the robustness of our data, SNPs with extremely low quality are filtered 690 out. Then the sequencing error rate, MCC rate and haplotype status were 691 carefully estimated. Subsequently, we calculated the likelihood of 692 observing SNP data from the SCM samples within regions proximal to the 693 disease-causing mutation site. The calculation is based on the single-SNP 694 likelihood, which was attained from a binomial model using the measured 695 allelic depth (AD) values and estimated parameters, and the recombination 696 probability that can be obtained from the DECODE dataset or other 697 datasets containing the recombination fractions. See Extended data Fig. 8 698 and Supplementary Notes S2 for more details. 699

700

Recursively estimating the MCC rate and haplotype status. MCC rate
 was defined as the ratio of DNA fragments originating from the maternal

chromosome. Haplotype status, representing the true parental origin of 703 DNA at each measured SNP, is categorized into three distinct classes: 704 'Paternal Chromosome Only,' 'Maternal Chromosome Only,' or 'Both 705 Parental Chromosomes.' For instance, the designation 'Paternal 706 Chromosome Only' indicates that the DNA identified at a specific SNP 707 exclusively originates from the father. Initial MCC rate and haplotype 708 status were estimated, and these parameters were then iteratively and 709 jointly calibrated until convergence (See Supplementary Notes S2 for 710 details). 711

712

Determining the inherited chromosome. Unlike previous publications, where a fixed number of SNPs were predetermined, our methodology involves a sequential addition of SNPs. Starting from the disease-causing mutation site, we incrementally incorporate SNPs one by one, calculating the log-likelihood ratio for each subset. This iterative process yields a curve of log-likelihood ratio plotted against the physical distance of the terminal SNP.

720

Typically, the curve initiates at a point where the ordinate is in close 721 proximity to zero, gradually stabilizing into a plateau with the addition of 722 sufficient SNPs. The distinctive properties of this curve enable the 723 determination of the inherited chromosome. Subsequently, the confidence 724 level of our disease-carrying analysis is categorized into four classes: 725 Highly Confident, Moderately Confident, Possibly confident, 726 and Undetermined, based on the characteristics of the curve. See 727 Supplementary Notes S2 for more details. 728

729

730 Haplotype phasing and whole genome reconstructing of SCM

In many regions of the genome from the SCM samples, genetic information from paternal and maternal origins of the embryo was not both detected. This was due to allelic dropout resulting from scWGA bias or LFLoH. The conventional diploid assumption of the genome could not be applied to haplotyping and genotype imputation of SCM samples.

Therefore, we developed a Pedigree-Population-based Imputation with 736 Haploid Assumption (PPIHA) to reconstruct the whole genome 737 information of SCM. Specifically, we combined pre-phased parental 738 haplotype information and selected informative SNPs where at least one 739 parental origin haplotype could be confidently phased, to create the 740 haplotype scaffold of SCM (Extended Data Fig. 4). For example, for an 741 informative locus with paternal genotype being AB (allele A comes from 742 haplotype P1 and allele B from P2, based on the pre-phased parental 743 haplotype), the maternal genotype being AA, and the genotype called from 744 SCM being BB, it could infer that the B allele of this variant in SCM is 745 derived from haplotype P2, based on the Mendelian rule (see Extended data 746 Fig. 4 and Fig. 9). Maternal origin haplotype was not determined at this 747 site due to ADO bias or LFLoH feature in SCM. And vice versa, for the 748 locus with maternal, paternal and SCM genotypes being AB, AA and BB, 749 respectively, maternal haplotype of SCM could be inferred while paternally 750 genetic information unavailable. After examining all such kind of 751 informative variants from the sequenced SCM data, we constructed a 752 sparse haplotype scaffold with single parent of SCM (Fig. 3A and Extended 753 data Fig. 9). 754

Paternal or maternal haplotypes in the scaffold were first corrected if 755 continuous variants of the opposite haplotype were observed on both 756 nearby sides. This correction was applied when the probability of a double 757 crossover between the nearest variants with the opposite haplotype, 758 calculated as the square of their genetic distances multiplied by 0.0001, 759 was less than 0.005 for the maternal haplotype or 0.001 for the paternal 760 haplotype, according to chromosome interference theory ([34], Extended 761 Data Fig. 9). Then, with parental haplotype support and HMM strategy, 762 paternal and maternal haplotypes of other variants were independently 763 phased based on pre-determined haplotype scaffold of SCM (Extended 764 Data Fig. 9). We assumed that the state (P1 or P2 for paternal haplotypes 765 and M1 or M2 for maternal haplotypes) at variant t could be inferred by 766 SNPs in the scaffold from t-5 to t+5. The transition probability between 767 two SNPs was considered equivalent to their genetic distance multiplied 768

by 0.01. Joint probabilities of the two states of paternal or maternal 769 haplotypes at variant t were calculated using a forward-backward 770 algorithm based on the adjoining ten variants (from t-5 to t+5). The 771 haplotype of variant t was then assigned to the state with the higher 772 probability. If the joint probabilities of the two states (P1/P2 or M1/M2) 773 for a variant were equal, phasing of that variant failed (Extended Data Fig. 774 9). By organizing the haplotype states of all variants according to their 775 chromosomal positions, the pedigree-based haplotypes of SCMs were 776 reconstructed. Subsequently, the genotypes of SCM were imputed with the 777 genome information of the parents (Fig. 3A, Extended Data Fig. 9). 778

Following pedigree-based genome reconstruction, population-based
genotype imputation was conducted using the 1000 Genomes haplotype
reference panel with Minimac4 [35], resulting in the reconstruction of the
whole genome information of the embryo from the SCM samples (Fig. 3A).
Polygenic risk score calculation

We employed a PRS model of T2D including 128 SNPs from a previous study [31]. To validate the model and determine the PRS percentile (indicating relative disease risk) in the UK Biobank (UKB) cohort, SNPs whose genotype not available or failing to be imputed in UKB cohort were excluded, leaving a total of 114 SNPs. We then directly used the effect sizes from the original study to calculate the PRS, employing the following formula:

791
$$PRS = \sum_{i=1}^{114} \beta_i \times G_i$$

where G_i is the allelic dosage and β_i is the effect size of SNP *i*.

Using the T2D model and control information in UKB cohort, PRS percentile for T2D was obtained to evaluate the polygenic risk of T2D for each individual.

796

- 797
- 798

DATA AVAILABILITY

799 Access to anonymized patient data is subject to a data-sharing

agreement and protocol approval from the institutional review board

801	committee. The study-specific analyzable dataset is, therefore, not											
802	publicly available. All summary statistics are available in											
803	Supplementary Tables.											
804												
805	CODE AVAILABILITY											
806	Code for BASE-niPGT-M is available at https://github.com/Ge-lab-											
807	pku/BASE-niPGT-M											
808												
809	REFERENCES											
810	32. Abecasis, G.R. et al. Merlinrapid analysis of dense genetic maps using sparse											
811	gene flow trees. Nat Genet 30, 97-101 (2002).											
812	33. Lander, E.S. and Green, P. Construction of multulocus genetic linkage maps in											
813	humans. Proc Natl Acad Sci USA 84, 2363-2367 (1987).											
814	34. Housworth, E.A. and Stahl, F.W. Crossover interference in humans. Am J Hum											
815	<i>Genet</i> 73, 188-197 (2003).											
816	35. Howie, B. et al. Fast and accurate genotype imputation in genome-wide association											
817	studies through pre-phasing. <i>Nat Genet</i> 44, 955-959 (2012).											
010												
819	ACKNOWLEDGEMENTS											
820	We thank Long Gao, Yuhang Huan, Ye Li, Cuiping Pan, Xiaodan Shi,											
821	Wenjie Sun, Zhongwei Wang, Liying Yan, Jingyi Yang, Jiakun Zhang											
822	and Nannan Zhang for their help. This work is funded by Beijing											
823	Advanced Innovation Center for Genomics at Peking University and											
824	Changping Laboratory. H.G. is supported by National Natural Science											
825	Fundation of China (No. 11971037 and T2225001). This study was											
826	also supported by the National Key R&D Program of China											
827	(2023YFC2705604), the National Science Foundation of China											
828	(82071721 and 82371706) and the special fund of the National											
829	Clinical Key Specialty Construction Program, P. R. China (2023)											
020												
030												

00	1
_ X.≾	
00	-

AUTHOR CONTRIBUTIONS

X.S.X., J.Q., L.H. and S.L. conceived and supervised the research. J.H. 832 and J.J collected SCM samples and performed PGT-M experiments of 833 biopsy samples. L.H. designed and performed the niPGT-M 834 experiments with the help from G.Y., D.C., Y.T. and K.L.. R.Z. and 835 H.G. performed the lineage analysis for niPGT-M with the help from 836 L.H., Q.W., Z.W., W.S. and X.C. Y.Z. and Y.X. performed the data 837 analysis for niPGT-P with the help from L.H. and G.C.. L.H., H.G., 838 Y.X. and Y.Z. wrote the initial manuscript with the help from Q.W. 839 and G.Y. All authors approved the final version of the manuscript. 840

- 841
- 842

COMPETING INTERESTS

L.H., H.G., R.Z. and X.S.X. are coinventors on patent application
PCT/CN2023/125605 that includes the experimental discovery and
BASE-niPGT-M in this manuscript. S.L., Y.Z. and Y.X. are current
employees of Yikon Genomics.

- 847
- 848

FIGURE CAPTIONS:

849

850 Figure 1 niPGT-M workflow using SCM samples.

A. Haplotype pre-phasing of parents is conducted using sequencing results 851 from the proband, grandparents, or at least a discarded embryo within the 852 family. Fertilized embryos are cultured, and immediately after transferring 853 the blastocysts to another dish for biopsy, SCMs are removed from each of 854 the residual drops. A few trophectoderm cells from each hatching 855 blastocyst were biopsied and sent to an external PGT lab for PGT-M testing, 856 serving as the gold standard. All the SCM samples underwent scWGA and 857 NGS, followed by SNP calling. 858

B. The SNP data of SCM samples, together with the phased haplotypes of the parents, are input into the BASE-niPGT-M method. After iteratively

estimating the MCC rate and haplotype states of each SNP, the loglikelihood ratio of inheriting disease-carrying chromosomes versus disease-free chromosomes was calculated. niPGT-M diagnose is performed based on the characteristics of the log-likelihood ratio curve as a function of distance to the disease-causing mutation site.

866

Figure 2 An illustrative case for our niPGT-M method.

- A. The SNP distribution and the linkage-analysis results of cfDNA in an
 illustrative case. This case involved an autosomal dominant genetic
 disease, where the husband carried a genetic variant for Marfan
 syndrome, a heritable connective tissue disorder. All eight SCM
 samples were collected, and cfDNA was amplified using the improved
 LIANTI method.
- **B.** The curves of log-likelihood ratio and diagnosis results using BASE-874 niPGT-M. The diagnosis results of five SCM samples resulted in a 875 classification of 'Highly confident', while one SCM sample was 876 classified into 'Moderately confident' due to our conservative diagnosis 877 strategy. The remaining two SCM samples were classified into 878 'Undetermined' due to either a very low density of SNPs or an opposite 879 inference of chromosome upstream and downstream of the disease-880 carrying mutation site. All these results were consistent with those of 881 trophectoderm biopsies, demonstrating no misclassifications. 882
- 883

Figure 3 Whole genome reconstruction of the embryo from SCM data.

Schematic illustration of the procedure for whole genome A. 885 reconstruction. This involves four steps. Firstly, parental haplotypes were 886 phased with sibling embryos or other family members. Secondly, using 887 pre-phased parental haplotype information and selected informative SNPs, 888 haplotype scaffold of SCM was created. Thirdly, pedigree-based 889 imputation was conducted on SCM with the genome information of parents. 890 Finally, whole genome information of SCM was reconstructed based on 891 population-based genotype imputation. 892

893 **B.** Whole-genome SNP density and T2D related SNPs coverage of 10T-

894 SCM1. The blue triangles were SNPs in PRS model of T2D directly called 895 from raw data; the orange triangles were SNPs imputed from pedigree 896 information; the gray triangles represent SNPs imputed from population 897 information.

- C. Haplotypes of SCM samples with different MCC and corresponding TE
 biopsy samples. Different haplotypes were showed with different colors.
 The uncolored areas denoted regions where reconstruction failed.
- 901 **D.** Whole-genome coverage of SCM samples before and after our 902 construction. See Supplementary Table S6 for the data.
- E. Haplotype and genotype concordances between SCM samples and the
 corresponding discarded embryos or TE biopsy samples. See
 Supplementary Table S6 for the data.
- 906

Figure 4 T2D associated genotype imputation and polygenic risk score calculation.

A. Distribution of reconstructed genotypes of T2D associated SNPs in SCM samples. T2D related genotypes in SCM were obtained from direct detection by NGS in cfDNA from SCM (blue bar), pedigree-based genotype imputation (green bar) and population-based genotype imputation (yellow bar).

- B. Genotype concordance of T2D associated SNPs in SCM versus
 corresponding TE or discarded embryo samples. Genotype 0/0, 1/1 and 0/1
 respectively denote homozygous reference allele, homozygous alternate
 allele and heterozygous genotypes.
- 918 **C.** PRS consistency of T2D between SCM and paired TE/discarded 919 embryo or amniotic fluid samples.
- 920 **D.** PRS percentile of T2D between SCM and paired TE/discarded embryo 921 or amniotic fluid samples. 20 SCM samples matched with TE samples 922 (green round dots), 5 SCM samples matched with discarded embryo 923 samples (blue rhombic dots), and 2 of the 20 TE biopsied embryos had 924 corresponding amniotic fluid samples, also be compared with the 925 corresponding SCM samples (red triangular dots).
- 926 See Supplementary Table S7 for the data shown in this figure.

927

Table 1. Summarized results of BASE-niPGT-M using all matched 928 biopsy/disgarded embryo. We categorized the results into three groups: 929 autosomal dominant disease, autosomal recessive disease, and X-linked 930 recessive disease. For autosomal recessive diseases, we compared each 931 allele of the embryos. The acronyms HC, MC, PO, UN, and RO stand for 932 Confident, Moderately Confident, Highly Possibly Confident, 933 Undetermined, and Ruled Out, respectively. Overall, we successfully 934 detected 95 out of 126 alleles, achieving a 75.4% average report rate. All 935 detected results were concordant with the gold standard, demonstrating a 936 100% accuracy rate of BASE-niPGT-M. See Supplementary Table S5 for 937 more detailed data. 938

939







Undetermined

В

Undetermined

High

High

ihood	pos	likelihood	pos	likelihood	pos	likelihood	pos	likelihood
.99	-2964.32	9.95	-450.45	23.58	-602.74	-9.6 -112.5		-35.55
.04	-1795.78	9.95	-291.82	25.0	-599.37	-9.26 -111.21		-34.16
.01	-1763.29	9.95	-168.07	20.0	-585.34	-8.73 -104.7		-30.0
.05	-1197.45	9.95	-124.7	19.23	-529.42	-4.61 -87.11		-27.13
.0	-1186.05	9.95	-114.32	19.23	-505.43	-7.48 -87.09		-26.91
.12	-1034.15	9.95	-113.5	15.0	-451.75	-4.74 -87.09		-23.49
.32	-783.49	9.95	-113.45	11.75	-451.28	-0.72 -68.91		-18.52
95	-672.38	9.79	-113.29	11.66	-441.04	-4.74 -49.81		-15.0
.11	-654.37	8.27	-88.34	9.04	-427.53	-0.73 -39.18		-10.0
.69	-130.97	4.16	-40.27	4.11	-145.28	3.36 -32.48		-5.0
.3	140.36	5.0	528.71	5.0	285.67	-5.0 77.28		-5.0
.72	289.75	10.0	581.86	5.15	314.93	-10.0 282.54		-10.0
0.1	442.72	6.44	582.58	7.07	322.24	-15.0 285.67		-13.78
0.36	560.06	9.99	629.95	11.57	337.32	-17.93 287.65		-13.82
0.36	587.64	12.58	631.35	13.4	338.94	-17.93 314.93		-15.0
1.44	602.12	12.85	756.08	13.44	376.42	-17.93 376.42		-17.99
1.96	704.09	12.87	761.86	13.44	522.13	-17.93 442.18		-17.99
5.69	767.06	12.88	881.83	13.44	597.71	-17.93 536.69		-17.99
8.76	813.73	12.89	957.82	13.44	649.12	-18.19 597.71		-18.13
5.69	814.02	12.89	957.89	13.47	688.69	-18.26 605.31		-18.19

High

High

Moderate

High



medRxiv preprint doi: https://doi.org/10.1101/2024.11.10.24317049; Dis version posted November 13, 2024. The copyright holder for this preprint (mich.was.not perijitied b per









Type of Disease	Family Number	Parent	# of Samples	# of Reportable Samples	# of Correct Samples	HC	MC	PO	UN	RO	Report Rate(%)
	1	М	6	2	2	2	0	0	0	4	33.3%
	2	Р	8	6	6	5	1	0	2	0	75.0%
Autosomal	3	Р	8	7	7	5	0	2	0	1	87.5%
Dominant	4	М	11	9	9	3	6	0	0	2	81.8%
Disorder	5	Р	4	3	3	1	2	0	0	1	75.0%
	6	Р	4	4	4	2	0	2	0	0	100.0%
	7	Р	4	3	3	1	0	2	1	0	75.0%
Total	\	\	45	34	34	19	9	6	3	8	75.6%
P tota		\	28	23	23	14	3	6	3	2	82.1%
M tota	nl in the second se	\	17	11	11	5	6	0	3	6	64.7%
	g	Р	5	3	3	2	1	0	0	2	60.0%
	0	М	5	2	2	1	1	0	1	2	40.0%
	Q	Р	13	9	9	5	2	2	2	2	69.2%
Autosomal	5	M	13	10	10	6	2	2	1	2	76.9%
Recesive	10	Р	5	3	3	2	1	0	1	1	60.0%
Disorder	10	M	5	4	4	3	0	1	0	1	80.0%
Disoraci	11	Р	8	7	7	2	4	1	1	0	87.5%
		М	8	7	7	4	2	1	1	0	87.5%
	12	Р	5	3	3	3	0	0	1	1	60.0%
	12	M	5	4	4	1	1	2	0	1	80.0%
Total alleles	\	\	72	52	52	29	14	9	8	12	72.2%
P tota		\	36	25	25	14	8	3	5	6	69.4%
M tota	<u>l</u>	\backslash	36	27	27	15	6	6	3	6	75.0%
X-Reccessive	13	М	6	5	5	5	0	0	1	0	83.3%
Disorder	14	М	3	3	3	3	0	0	0	0	100.0%
Total/ M total	\	\backslash	9	8	8	8	0	0	1	0	88.9%
Tot		126	94	94	56	23	15	12	20	74.6%	
P tota			64	48	48	28	11	9	8	8	75.0%
M tota		62	46	46	28	12	6	7	12	74.2%	

940

941

ADDITIONAL INFORMATION

- **Extended** data 942
- 943

Extended Data Fig. 1 DNA size distribution in Spent Embryo Culture 944 Medium (SCM) 945

Size distributions of DNA molecules in six spent culture medium samples 946 with different DNA fractions (Fragment Analyzer System). The horizontal 947 axis represents the range of DNA fragments detected. 1bp is the lower 948 marker, and 6000bp is the upper marker. The vertical axis represents a unit 949 of relative fluorescence units. The first peak of insert fragments was 950 approximately 180bp, with subsequent peaks spaced at 180bp intervals, 951 corresponding to the length of DNA wrapped around a nucleosome. SCM 952 2, SCM4, SCM5 and SCM6 showed varying levels of apoptosis, 953 suggesting different degrees of DNA degradation. Additionally, SCM3 954 failed to produce an effective library due to its low cfDNA content, and 955 SCM1 did not exhibit the characteristic apoptosis signals. See 956 Supplementary Notes S1 for more experimental details. 957

958

Extended Data Fig. 2 Genome Coverage of Spent Embryo Culture 959 Media 960

Genome coverage of all SCM samples. 961

962

Extended Data Fig. 3 Illustration of large fragment loss of haplotype 963 (LFLoH) in one sample of SCM. 964

Blue and orange bars denote haplotypes respectively from paternal and 965 maternal chromosome 1. Black region represents no genetic information 966 observed in specific haplotypes. 967

968

Extended Data Fig. 4 Strategy for informative SNPs selection. 969

0 and 1 respectively denote reference and alternative allele of variant. All 970

the informative SNPs selected are listed. 971

972

973 Extended Data Fig. 5 Estimated MCC rates.

The top and bottom panels illustrate the MCC rates for autosomal dominant and autosomal recessive family samples, respectively. Circles represent pathogenic loci inherited paternally, while triangles denote maternal inheritance. Hollow symbols indicate the original MCC rates, and solid symbols represent the final MCC rates. Blue symbols correspond to diagnosable samples, whereas red symbols denote undiagnosable samples. See Supplementary Table S4 for the data.

981

982 Extended Data Fig. 6 Improved protocol from LIANTI Method.

In this study, we retained the LIANTI single-cell amplification method, 983 which involves (a) : 1) Transposon-assisted DNA fragmentation; 2) Gap-984 filling and extension; 3) Reverse transcription; 4) Second-strand synthesis; 985 5) Library construction and sequencing. To adapt this method for culture 986 medium samples, we made the following improvements (b): 1) Adjusted 987 the lysis system with a 10.8 µl sample volume and increased QIAGEN 988 Protease (Qp) concentration to 35 mg/ml; 2) Doubled the transposon 989 concentration; 3) Added a 19bp transposon ME sequence complementary 990 primer to enhance reverse transcription efficiency; 4) Included 10 991 additional PCR cycles in second-strand synthesis to improve the 992 conversion of single-stranded cDNA to double-stranded DNA. 993

994

995 Extended Data Fig. 7 Improved LIANTI amplification is more suitable 996 for amplifying culture medium samples.

Clinical collection of SCM (spent culture medium) yields limited volumes, 997 and samples are precious. In this study, we utilized the BJ cell line to 998 simulate the embryo culture process. a) Size distributions of DNA 999 molecules in BJ15 and BJ17 with different DNA fractions (Fragment 1000 Analyzer System). b) Genome coverage analysis of BJ15 and BJ17, which 1001 were amplified using Improved LIANTI, MALBAC, and MDA. c) SNP 1002 number analysis of BJ15 and BJ17, which were amplified using Improved 1003 LIANTI, MALBAC, and MDA. Genome coverage and SNP number 1004

normalization were carried out with Bulk sample. The sequencing depth
for bulk genomic DNA (gDNA) exceeds 30X, while for amplified products,
the sequencing depth exceeds 10X. BJ15: BJ cell lines medium collected
on Day5; BJ17: BJ cell lines medium collected on Day7 (see
Supplementary Table S2).

1010

1011 Extended Data Fig. 8 The workflow of BASE-niPGT-M.

The Bayesian model uses the sequencing data of SCM and the phased 1012 haplotypes of the parents as input. By iteratively estimating the MCC rate 1013 and haplotype states of each SNP, it calculates the log-likelihood ratio of 1014 inheriting disease-carrying chromosomes versus disease-free 1015 chromosomes as a function of distance to the disease-causing mutation site. 1016 This curve, with its distinct characteristics, allows classification of the 1017 results into four categories: Highly Confident, Moderately Confident, 1018 Possibly confident, and Undetermined. 1019

1020

1021 Extended Data Fig. 9 An example illustrating the workflow of 1022 pedigree-based whole genome reconstruction.

In the initial called genotype of the SCM, red numbers indicate the 1023 informative alleles, which can identify the inherited allele from one parent. 1024 Black regions represent the LFLoH pattern, characterized by consecutive 1025 SNPs missing the same haplotype. Blue letters denote the imputed 1026 haplotypes on non-informative SNPs, determined using the HMM method 1027 with genetic distance as the transition probability. Green letters indicate 1028 corrected haplotypes from the raw phasing results by the HMM method, 1029 with incorrect ones shown in red. Green numbers show the imputed allele 1030 or genotype of the genome. 1031

1032





perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

СМ

ΤE



Paternal genotype	Maternal genotype	SCM genotype	Key allele	SNP type
0/1	0/0	0/1	1	Paternal informative SNP
0/1	0/0	1/1	1	Paternal informative SNP
<mark>0</mark> /1	1/1	<mark>0</mark> /1	0	Paternal informative SNP
<mark>0</mark> /1	1/1	0/0	0	Paternal informative SNP
0/0	0/1	0/1	1	Maternal informative SNP
0/0	0/1	1/1	1	Maternal informative SNP
1/1	<mark>0</mark> /1	<mark>0</mark> /1	0	Maternal informative SNP
1/1	<mark>0</mark> /1	0/0	0	Maternal informative SNP





5'-CTGTCTCTTATACACATCTGAACAGAATTTAATACGACTCACTATAGGGAGATGTGTATAAGAGACAG-3' 5'-CTGTCTCTTATACACATCTGAACAGAATTTAATACGACTCACTATAGGGAGATGTGTATAAGAGACAG-3'



0.5ul 35mg/ml QP			·	
	 <i>.</i>			
š				

а



BASEni-PGT-M

DENIFITIONS

10MB (Million Base pairs) upstream and downstream of the disease-causing mutation site in the biological sample is defined as the adjacent region of the disease-causing mutation site.

Maternal Cell Contamination rate (MCC rate) is defined as the proportion of maternal DNA in the culture medium of an embryo compared to the total amount of DNA.

Large Fragment Loss of Haplotypes (LFLoH) is defined as the phenomenon that long and contiguous DNA fragments from one or both parents do not enter the culture medium, and the hyplotype status of each SNP is defined to represent whether only DNA from one parent is present in the culture medium.

Amplify the DNA molecules from spent embryo culture medium, sequencing and performing mapping and SNP calling, filtering the SNPs of extremely low quality.

INPUT

Initially estimate the SER, the relative density of SNPs, the MCC rate and the hyplotype status of each SNP.

Recursively estimate and update the MCC rate and the haplotype status of each SNP, using a conservative strategy with the single-SNP likelihood.

medRxiv preprint doi: https://doi.org/10.110//2024.11.10.24317049; this version posted November 13, 2024, preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license. It is made available under a CC-BY-NC-ND 4.0 International license. It is made available under a CC-BY-NC-ND 4.0 International license. data within regions adjacent to the disease-causing mutation site, combining the likelihood at each single SNP and the recombination probabilities.

$$P(AD_{n:end} | F_n = i, M_n = j) = L_{ij}^{(n)} \sum_{k,l=1}^{2} P(AD_{n+1:end} | F_{n+1} = k, M_{n+1} = l) P(F_{n+1} = k, M_{n+1} = l | F_n = i, M_n = j)$$

AD is the measured Allele Depth data of SNPs.



Conduct haplotype pre-phasing, for the blood samples of both parents or solely the disease-carrying parent.

Incorporate the maternal cell contamination (MCC) rate and haplotype status into the likelihood of the four possible inheritance scenarios at each single SNP representing whether the genetic chains are inherited from the paternal or maternal alleles of the parents.

$$\hat{L}_{ij}^{(n)} = \frac{AD_n!}{AD_{0,n}! \cdot AD_{1,n}!} \Big[\hat{p}_{ij}^{(n)} \widehat{SER} + \Big(1 - \hat{p}_{ij}^{(n)}\Big) \Big(1 - \widehat{SER}\Big) \Big]^{AD_{0,n}} \\ \cdot \Big[\hat{p}_{ij}^{(n)} \Big(1 - \widehat{SER}\Big) + \Big(1 - \hat{p}_{ij}^{(n)}\Big) \widehat{SER} \Big]^{AD_{1,n}}$$

SER is the estimated sequencing error rate, and $\hat{p}_{ij}^{(n)}$ is the theoretical probability that an ALT gene is obtained by randomly sampling one read at n-th SNP.

Recombination probability derived from database

$$P(F_{n+1} = k, M_{n+1} = l | F_n = i, M_n = j)$$

The linkage state of the n-th SNP: $F_n=1(P1), F_n=2(P2), M_n=1 \text{ (M1)}, M_n=2(M2).$



Phasing informative SNPs of SCM								Building haplotype of SCM Reconstructing genotype			be of SCM		
Pate	ernal	Mat	ernal	Initial		Paternal	Maternal	Comotio	Paternal	Maternal	Paternal	Maternal	Imputed
haplo	otype	haple	otype	genotype of	SNP type	Haplotype of	Haplotype of	distance	Haplotype of	Haplotype of	allele of	allele of	genotype of
P1	P2	M1	M2	SCM		SCM	SCM	uistance	SCM	SCM	SCM	SCM	SCM
1	0	0	0	1/1	Haploid	P1		0.12	P1	M1	1	0	1 0
1	1	0	1	0/0	Haploid		M1	0.48	P1	M1	1	0	1 0
0	0	1	0	1/1	Haploid		M1	0.61	P1	M1	0	1	0 1
0	1	1	1	0/0	Haploid	P1		0.85	P1	M1	0	1	0 1
1	1	0	1	1/1	Uncertain			1.02	P1	M1	1	0	1 0
0	1	1	1	0/0	Haploid	P1		1.16	P1	M1	0	1	0 1
1	0	1	1	0/0	Haploid	P2		1.22	P1	M1	1	1	1 1
0	1	1	1	0/0	Haploid	P1		1.34	P1	M1	0	1	0 1
1	1	0	1	<mark>0</mark> /1	Diploid		M1	1.98	P1	M1	1	0	1 0
1	0	0	0	0/1	Diploid	P1		2.39	P1	M1	1	0	1 0
0	0	1	0	1/1	Haploid		M1	2.64	P1	M1	0	1	0 1
1	1	1	0	0/0	Haploid		M2	2.73	P1	M1	1	1	1 1
0	0	1	0	1/1	Haploid		M1	2.85	P1	M1	0	1	0 1
0	0	1	0	1/1	Haploid			2.96	P1	M1	0	1	0 1
1	0	0	0	0/1	Diploid	P1		3.15	P1	M1	1	0	1 0
0	0	1	0	0/1	Diploid		M1	3.21	P1	M1	0	1	0 1
0	0	0	1	0/1	Diploid		M2	4.37	P1	M2	0	1	0 1
1	0	0	0	0/1	Diploid	P1		4.93	P1	M2	1	0	1 0
0	1	1	0	0/1	Diploid			5.30	?	M2	?	0	./.
1	0	1	1	<mark>0</mark> /1	Diploid	P2		5.67	P2	M2	0	1	0 1
0	1	0	0	0/1	Diploid	P2		5.69	P2	M2	1	0	1 0
0	0	0	1	1/1	Haploid		M2	5.74	P2	M2	0	1	0 1
0	1	0	0	1/1	Haploid	P2		5.88	P2	M2	1	0	1 0

Supplementary Information 1033

1034

- **Supplementary Notes** 1035
- 1036
- **Supplementary Tables** 1037

1038