# nature research

Corresponding author(s):	Yazhuo Zhang
Last updated by author(s):	Nov 4, 2020

# **Reporting Summary**

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our <u>Editorial Policies</u> and the <u>Editorial Policy Checklist</u>.

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section

_					
5	۲a	ŤΙ	ıct	ш	<u>ر</u> د

	an statistical analyses, commit that the following frems are present in the figure regent, main coxt, or methods section.
n/a	Confirmed
	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
	The statistical test(s) used AND whether they are one- or two-sided Only common tests should be described solely by name; describe more complex techniques in the Methods section.
	X A description of all covariates tested
	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
	For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i> ) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
×	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
	Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i> ), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

### Software and code

Policy information about <u>availability of computer code</u>

Data collection

Patient clinical and demographic data was collected using Microsoft Excel (V16), which was added in the manuscript.

Data analysis

All computer softwares used for bioinformatical and statistical analyses in this study were described in Methods. Maftools (v1.6.05), SigProfiler (v0.0.5.77), Meerkat (v0.189), dNdScv ( HYPERLINK "https://github.com/im3sanger/dndscv/releases/tag/0.1.0" v0.1.0), pyclone (v0.13.1), facets (v0.5.6), ShatterSeek (v0.4), PennCNV-seq (v1.0.4), Canvas (v1.31), BWA (VN:0.7.8-r455), GATK Picard (v2.18), MoCCA-SV (V0.2), Svaba (v1.1.0), Breakdancer (v1.4.5), Delly (v0.8.1), Manta (v1.4.0), BAM-matcher (2016 version), sentieon (sentieon-genomics-201711.02), SomaticSeq (v2.7.2),

IGV (v2.3.61), GATK HaplotyperCaller, UnifiedGenotyper LeftAlignAndTrimVariants (v3.8).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

### Data

Policy information about availability of data

All manuscripts must include a <u>data availability statement</u>. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Whole genome sequencing data can be accessed through GSA (https://bigd.big.ac.cn/gsa/) under Accession# PRJCA002098). All other relevant data are available on request from the corresponding author (YZ). We have added the data availability statement in the manuscript with the accession number. We have also added the download site for TCGA data.

t 🖆 Ălś lś 🖆 ÖÖÜŚ 🗖	má ĽámajŎ ÖVÁÖN ÖVÁ ĽálÖzeÚn Tegî 'n ráláŘrí Na Líz Tegî Řrá men Ölî ráj ráðir ÖVÁ ŘAJA neu Jan MÁĞ lái ÖNHenl Lázenrá (ﷺA∰NHA † Tegî n lámáir
× [去 lǐ去∎ǐśl	ŚႷĂŌႸij'nĂϼϼჼჇႼჼႷႽႼჼśჇ9º¤ϼϼϯ┩ĂϼŚŌϼϼĠႩႼĂ╓ŦŹŚΦŌႸ┲₽∰ŚŧŒČĂϼჼႷႸႼჼŚ₽
OminĂinśźśinś∎ľśľou,∓oźÓ	Ÿśŕ¤ľĵ∰ś∎ĊŎŶŸĂĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸ
r lks ov lk_	_P ム O O か な 丁 な ム O III
[ <del>2</del> S (1 <del>-</del> S	uľśłłĠŕŦŕśł <mark>┦</mark> ■
! <b>ਜ਼</b> Ġŕቴŧ <b>█</b> ĵŧ♂ŧ	i'œqué œw Ölélé Jæluû éōéw Öléw Ölé f√liœquî iné√k wé-¦ĂÖbé;
{Ā <b>∰.4</b> ≦ ℓ-tiś	ا كَالْمُهُمُ اللَّهُ اللّ
	zăndis ănistă D( śzarczani "laiń a A Clistă Jattie Oăt nactordia ătristă librii pădisi liboăt ristină librii libri
	ĬĂĦĬĠŊŖĠŧĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸĸ
	『中年前 置比的 no And
EĂĠĂ ÁŖPŶ DLL	bor ŘÁŠOŘI sílôr žido Čls řoši
5ĂĊĂŚŜľ≟ałł	
wé Jada ACha	Ç╙╙╙╍òҲӆӆ┩Ӑѵ҈ҕ҆ѯҥҧӱҡҾӷ҅ҭӷ╙ҧҕ҈Ӄӿҧ҈҃ѕӷӹӱҕҹ҇Ҹ҈҅Ѱ╍ѻ҅҈ҪӢҋӡӷѻӹӷ҅ӷ҅҅҅҇Ҫ҅ӷҡҕӽӗӼ҈ӷ҈ӷӷҕӷҸ҇ҋҕӈ҅Ӓ҈Ҫ҅҈҈ҧҥ҅Ӧ҈ѩҪ҅ ┇҈ӷ҇҇ѽ҅Ҫ҅҅҅҅҅҅҅҅҅҅҅҅҅҅҅҅҅҅҅҅҅҅҅҅҅҅҅҅҅҅҅҅҅҅
	<del>-</del>
wăsí si tăčis	ÇHÉ ÇÎ FÎ TÎ FÊ ÇÎ HÎ SÊ TÎ BÎ FÊ ÎNÊ ÎNÊ ÎNÊ ÎNÊ ÎNÊ ÎNÊ ÎNÊ ÎNÊ ÎNÊ ÎN
. <b>"</b> ##ŕ <del>***</del> +	Çlé (Çri T-litis the fight of
<u> </u>	

í śrájj) tá táng Áðbo áro já dong Álog Óto sátask og jágrafu sátask síta sátask ár sátask í sátask í sátask í s Læg on sátor tást trá áða óða tog ngát títag árá og jásta lóts játask ár tog narksát trá ár átarotaláská láiðbo látar kást de tástaská

a ĂĠ'nĂ₄³ śŝ√ś'n∰ś∎Œ́→▞ૠĠ∰ℓ	a śĊ <b>l</b> ∎ŕł
■□Ă Luōunāśŕ ¼ ÖŚţţřŦ	■A Laōaasí 4 Ölsł¢† Ŧ
x	x / Ut Réij
x ☐ 9j ∰rnīmaĠ l'śpodeśł	X God i tage
▼ t Ăăśo o co a † T Ă o a † T	ː a wl⊞'Al'sr ∎sĵ 'n∎∰A <del>  l  </del>
▼ ! ■ ĂĂĂĂ ĂŒſ ŒĊĬŚ'nŒ'nŊĂŒ₹	
☐ X lĵ∰Ă∎'nśłśĂn" LjĂnĠHjĂ∎Ġ	
x / H A A A A	
5ĵĂ iś'nsłśĂn Loźlowiś'n	

# Iĵ∰Ă∎'nsłśÄn" LJÄnÖHJAŒŮ

[ المسلم المراقية المسلم المراقية الم

wsí''nĵ-√ooosecc

9Ö441 **□**ŌŚr12441C

ÇÜS (ÇÍT JANGALAĞA AJANAĞI LETÜS SÖÜL LAMAMA KÜSS AZÖÜS . SÜÜL ÇAMAĞALA ALAĞAĞA ÖĞ MAZAĞIĞI İLAMESIĞAĞA LAĞAĞI ALAĞMIŞI ANAMAÇÎÇIT JAĞÜLAĞI

boĠ ďĂĊţ — Lái ĂĠbo oo Ởś ĂIJnoōĂ pźŰś ŁĢ ŕ Ŧ "noĠl o " Î ŁĊĂŁo Łś "noōł śŕ - lo Űś " Ăoj Łľnlæ

## Dual use research of concern

Alter the host range of a pathogen

Enable evasion of diagnostic/detection modalities

Enable the weaponization of a biological agent or toxin

Any other potentially harmful combination of experiments and agents

Policy information about <u>dual use research of concern</u>

### Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:
No Yes    Public health     National security     Crops and/or livestock     Ecosystems
Any other significant area  Experiments of concern
Does the work involve any of these experiments of concern:
No Yes  Demonstrate how to render a vaccine ineffective  Confer resistance to therapeutically useful antibiotics or antiviral agents  Enhance the virulence of a pathogen or render a nonpathogen virulent
X     Increase transmissibility of a nathogen