



石乐明 博士

复旦大学生命科学学院 教授

复旦大学附属肿瘤医院 双聘教授

深圳微芯生物科技有限公司 共同创始人

国家重点研发计划项目 首席科学家

国家特聘专家(中组部千人计划入选者)

从事精准医学、药物基因组学和医学大数据研究，旨在提高新药研发成功率和药物临床使用有效率

发表论文200多篇(含15篇Nature Biotechnology)，SCI他引13,000多次，h-index 55；主导或参与制定了4项ISO和CLSI组学标准

曾任美国FDA资深研究员12年，发起国际组学大数据质量控制联盟MAQC (2004)，自然出版集团为其出版了4个专辑

共同创立国际MAQC组学大数据质量控制学会(2017)，旨在提高多组学技术的可靠性，担任首任主席(2017-2018)

长期开展组学大数据的质量控制与标准化研究

包括数据产生、数据分析、数据解读

- Data generation >16 years
- Data analysis >100 organizations
- Data interpretation >500 participants



FDA 为人类表型组计划的实施奠定了质控技术和国际合作的基础

www.nature.com/nbl/focus/maqc/ (2006)
www.nature.com/focus/maqc2/ (2010)
www.nature.com/nbl/collections/seqc/ (2014)
Nature Biotechnology (September 3, 2021)



WG#1: Somatic Mutation (WGS/WES)

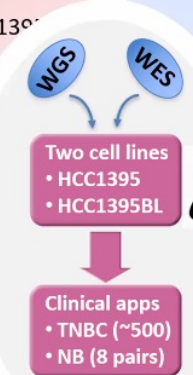
- Cross-lab and -platform analysis
- Additional samples
 - Xenograft/FFPE (HCC1395)
 - Spike-in with Sequins

WG#2: Target Gene Seq (TGS)

- Cross-lab and -platform analysis
- Additional samples
 - 10 Agilent UHRR cell lines
 - Sample B (normal) and Bm (Homozygous)

WG#3: Germline Mutation

- Joint effect of pipeline and coverage
- Additional datasets:
 - NCTR TRIO and GIAB
 - Chinese Quartet
 - ABRF data using GIAB



WG#4: Difficult Genes

- Sensitivity/specificity
- Additional datasets:
 - NCTR TRIO and GIAB
 - Chinese Quartet
 - ABRF data using GIAB

WG#5: Genome Assembly

- Personal genome vs reference-based approaches
- Additional datasets:
 - NCTR TRIO and GIAB
 - Chinese Quartet
 - ABRF data using GIAB

WG#7: Epigenetics QC

- Benchmarking and platform
- Additional datasets

MAQC-IV (SEQC2): DNA-seq (Precision Medicine) 2014-2021??

Courtesy of Dr. Weida Tong, US FDA

Oncopanel Sequencing Working Group

RESEARCH Open Access

A verified genomic reference sample for assessing performance of cancer panels detecting small variants of low allele frequency

Wendell Jones¹, Brinberg Gong¹, Natalia Novikova¹, Dan Li¹, Rebecca Kuko¹, Todd A. Richmond¹, David J. Johnson¹, Hui Bai¹, Javed Mirza¹, David Mirza¹, David Mirza¹, James R. Bucher¹, Mehdi Roshan¹, Katherine Wilcox¹, Marco Cheng¹, Wenjun Bao¹, Lee Scott Beahm¹, Anne Bergstrom Lucas¹, Daniel Burgess¹, Daniel J. Burt¹, Simon Cassidy¹, Chang Jiang Chang¹, Guangchun Chen¹, Tian Chen¹, Yun-Cheng Chen¹, Daniel J. Clark¹, Angela Lee Fong¹, Jonathan Fow¹, Margherita Franceschini¹, Yuhua Fu¹, Cesare Fusco¹, Krishna Gaddam¹, Kira P. Gao¹, Majken Gao¹, Yingsi Hao¹, Scott Haggard¹, Gurgen Harizan¹, Nathan Harlow¹, Jeff Isler¹, Giuseppe Jarama¹, David Philip Keel¹, Pascal Lalonde¹, Anshu Li¹, Jianjun Li¹, Quan-Zhen Liu¹, Yifeng Liu¹, Zhiqiang Liu¹, Zhechao Liu¹, Maria Sofia Lopez¹, Kelli Mickus¹, Raymond Miller¹, Wenyi K. Mittal¹, Manoj Mohandas¹, Carlos Patino-Perez¹, Barbara L. Parsons¹, Fajun Qiu¹, Andrea Schmitt¹, Taku Sato¹, Kaye Steingard¹, Chen Sun¹, Masha Tsvetkova¹, Dong Wang¹, Zhong Wang¹, Leihong Wu¹, Weizhong Xue¹, Chang Xu¹, Ying Xu¹, Jianjun Zhang¹, Wan Zhang¹, Zhiqing Zhang¹, Haoting Zheng¹, Christopher E. Nelson¹, James C. Wiley¹, Weida Tong¹, Lening Shi¹ and Joshua Xu¹

ARTICLES

Evaluating the analytical validity of circulating tumor DNA sequencing assays for precision oncology

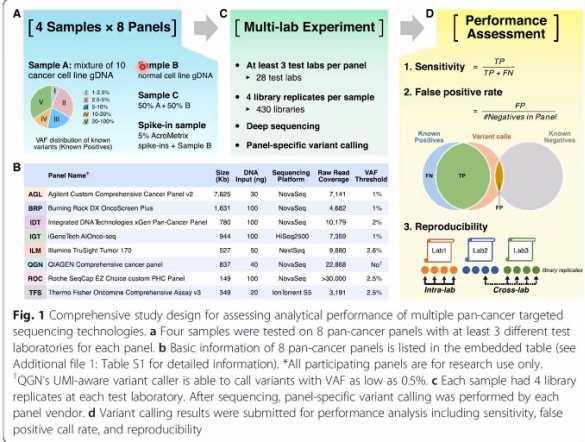
Dr. Joshua Xu (US FDA)

RESEARCH Open Access

Cross-oncopanel study reveals high sensitivity and accuracy with overall analytical performance depending on genomic regions

Brinberg Gong¹, Dan Li¹, Rebecca Kuko¹, Natalia Novikova¹, Wan Zhang¹, Zhang Wang¹, Carlos Patino-Perez¹, Zhong Wang¹, Kevin Li¹, Wanxin Cai¹, Jennifer S. LiCoro¹, Eric Lader¹, Todd A. Richmond¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark¹, Daniel J. Johnson¹, James C. Wiley¹, James R. Bucher¹, Ying-Cheng Chen¹, Guangchun Chen¹, Daniel Burgess¹, Simon Cassidy¹, Wenjun Bao¹, Nathan Goyal¹, Nathan Haggard¹, Fajun Qiu¹, Katherine Wilcox¹, Marco Cheng¹, Christopher E. Nelson¹, Daniel J. Burt¹, Jeffrey Conroy¹, Anne Bergstrom Lucas¹, Anne Bergstrom Lucas¹, Marica Brucato¹, Hui Bai¹, James Buchanan¹, Thomas M. Butler¹, James C. Wiley¹, James R. Bucher¹, Hui Bai¹, Masha Tsvetkova¹, Daniel J. Clark

Comprehensive Study Design for Assessing Analytical Performance of Eight Oncopanel



14

Summary of Recommendations for Oncopanel Sequencing

Issue	Recommendation
Reference material selection	A reference material (or a set of reference materials) with a high density of known variants spanning a range of low allele frequencies (e.g., from above 1 to 20%) is needed to assess the analytical performance of oncopanel (Fig. 1a). Ideally, the reference material will encompass a diversity of simple and complex variants in different cancer-associated genes.
Sensitivity	Sensitivity was found to be high (>96.5%) for variants previously verified to have variant allele frequency (VAF) greater than 5% (Fig. 2a). We recommend that a less stringent VAF threshold is used to achieve high sensitivity for variants of VAF below 5%.
Spike-in	Utilizing a sample spiked-in at a specified amount (e.g., 5%) can provide additional variants at known allele frequencies for analytical validation of oncopanels.
Variant type	The sensitivity for detecting insertion and deletion variants (indels) is typically more variable and poorer than single nucleotide variants (SNVs), and this difference becomes more pronounced at low VAFs (Fig. 3). We recommend that analytical validation of indels is performed independently to SNVs.
Determining VAF threshold	Reference materials can be used to establish an optimal VAF threshold that reduces false positives (FPs) and retains sensitivity. Indels and SNVs may require different VAF thresholds to optimize performance (Fig. 3b).
Controlling the FP rate	In applications where a minimal FP rate is required, raising the allele frequency threshold was effective at reducing FPs. The additional restriction of analysis to the consensus targeted regions (CTR) can further reduce FP rate.
Genomic location	Genomic location can impact the rate of FPs detected, and we recommend that analytical validation of panels is independently performed inside and outside of the CTR (Fig. 4a).
Cross-lab reproducibility	Measuring the analytical performance of a panel in multiple labs is critical to establish reproducibility (Fig. 4c).
Estimating tumor mutational burden (TMB)	TMB estimation should be confined to the CTR of each panel. Applying a minimal VAF threshold was helpful to reduce FPs and improve TMB evaluation (Fig. 5d).

15

Sensitivity, False Positives, and Reproducibility

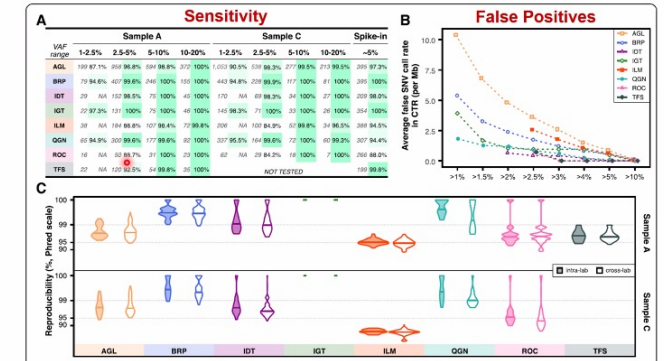
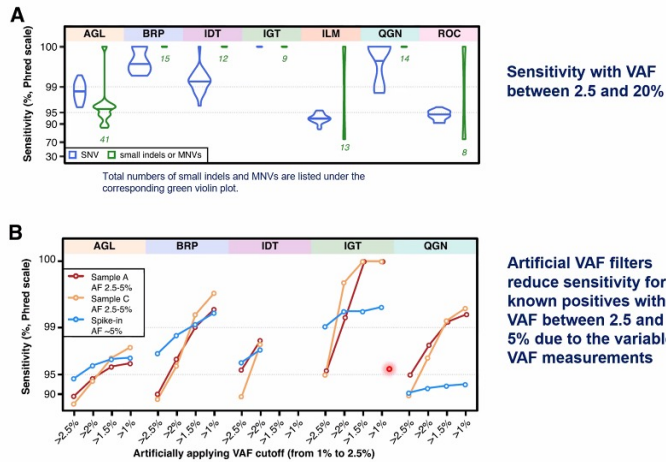


Fig. 2 Reproducibility and sensitivity across VAF ranges for SNVs in the consensus targeted regions. **a** Table listed the number of known variants in each VAF range (left number), sensitivity (right number) for all 8 panels across all samples tested. For the panels with a built-in VAF threshold, "N/A" is listed if the VAF low bound is much lower than the panel provider's chosen VAF threshold. The VAF threshold is 2.6% for ILM, 2.0% for IDT, 2.5% for ROC, and 2.5% for TFS, respectively. **b** Average false positive SNV calls per million across various VAF cutoffs. Jittering was applied to avoid overlapping. **c** Cross-lab and intra-lab reproducibility (in Phred scale) for variant calls with VAF between 2.5 and 20%.

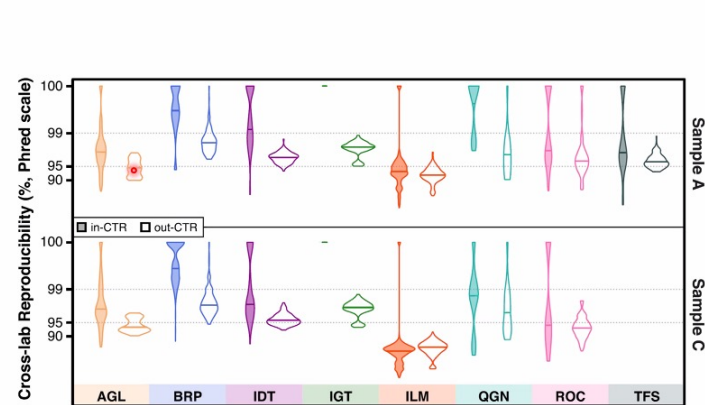
16

Impact of Variant Type and VAF Cutoff on Sensitivity



17

Cross-lab Reproducibility: Within CTR > Outside CTR



18

Technical, Intrinsic, and Overall CV in Estimating TMB

Large panel is needed for accurate TMB estimation

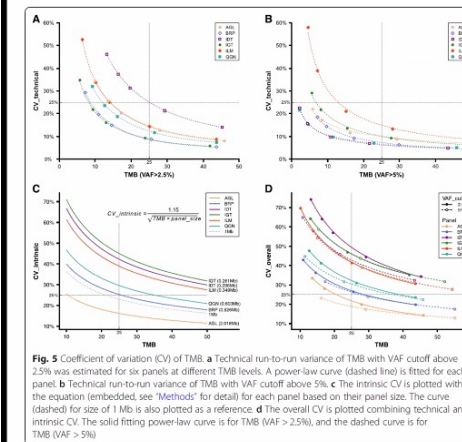


Fig. 5 Coefficient of variation (CV) of TMB. **a** Technical run-to-run variance of TMB with VAF cutoff above 2.5% was estimated for six panels at different TMB levels. A power-law curve (dashed line) is fitted for each panel. **b** Technical run-to-run variance of TMB with VAF cutoff above 5%. **c** The intrinsic CV is plotted with the equation (embedded; see "Methods" for detail) for each panel based on their panel size. The curve (dashed) for size of 1 Mb is also plotted as a reference. **d** The overall CV is plotted combining technical and intrinsic CV. The solid fitting power-law curve is for TMB (VAF > 2.5%), and the dashed curve is for TMB (VAF > 5%).

Technical variance < Intrinsic variance associated with estimating exome-level TMB from a smaller focused panel

Overall CV ~ Intrinsic CV

To maintain an overall CV < 25% in the clinically important TMB range (15–30), AGL (7,625 Mb, the largest panel) combined with an VAF cutoff at 5% would be the only panel from these eight suitable for TMB estimation

19

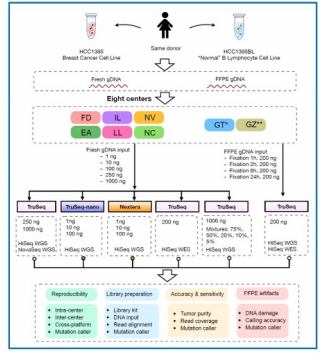
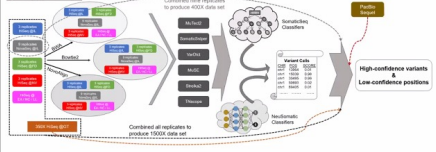
WG1: Somatic Mutation Working Group (WGS/WES)



Towards best practice in cancer mutation detection with whole-genome and whole-exome sequencing
 Wenming Xiao, 103
 Email: wenming.xiao@fda.hku.hk
 Xiao W, Nat 2

Articles
 Establishing consistency reference samples, data and call sets for benchmarking cancer mutation detection using whole-genome sequencing
 Nat Biotechnol 2019

A multicenter study benchmarking single-cell RNA sequencing technologies using reference samples
 Wang Q, Zhang H, Wang J, et al. Nat Biotechnol 2019



Tumor-normal paired DNA samples from a breast cancer cell line and a matched lymphoblastoid cell line enable calibration of clinical sequencing pipelines and benchmarking 'tumor-only' or 'matched tumor-normal' analyses.

实现三阴性乳腺癌的精准分型和精准用药, 提高患者生存

- 基于准确可靠的分子表型组数据, 实现三阴性乳腺癌的精准四分型
- 根据各分型的分子特征, 制订最佳治疗方案, 改变临床实践, 提高患者生存

Multi-omics profiling

468 Chinese TNBCs

Subtype-specific precision medicine

Subtype	Target	Drug	Response
BLB	ERBB2	Tyrosine kinase inhibitors	~50% ORR
LAR	AR	Androgen receptor inhibitors	~10% ORR
MES	CDK4/6	CDK4/6 inhibitors	~30% ORR

References:
 Jiang YZ et al., Cancer Cell 2019
 Xiao Y et al., Clinical Cancer Research 2019



- 建立“基因-突变-疾病-药物”四者相互关联的精准医学知识库系统与精准用药解读系统
- 累计生成5000多份乳腺癌多基因检测报告, 辅助乳腺癌患者精准用药

<http://www.fudan-pgx.org/premedkb/>

Yu Y et al., Nucleic Acids Res 2019

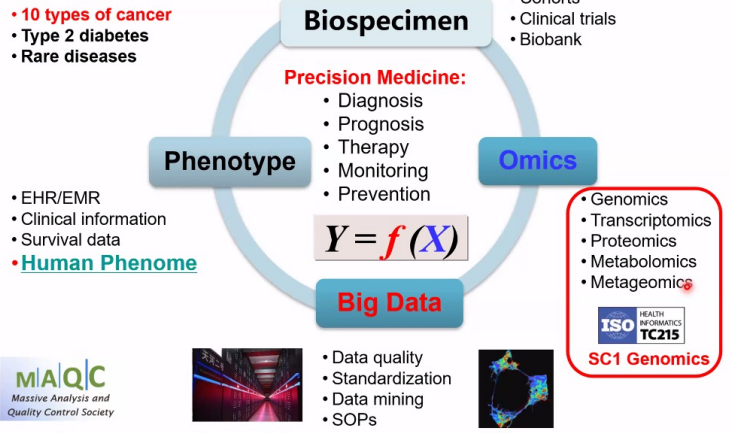
PreMedKB: an integrated precision medicine knowledgebase for integrating relationships between Genotype, phenotype, variation and drug

Lang GT et al., Nature Commun 2020

Characterization of the genomic landscape and actionable mutations in Chinese breast cancers by clinical sequencing

Toward Precision Medicine

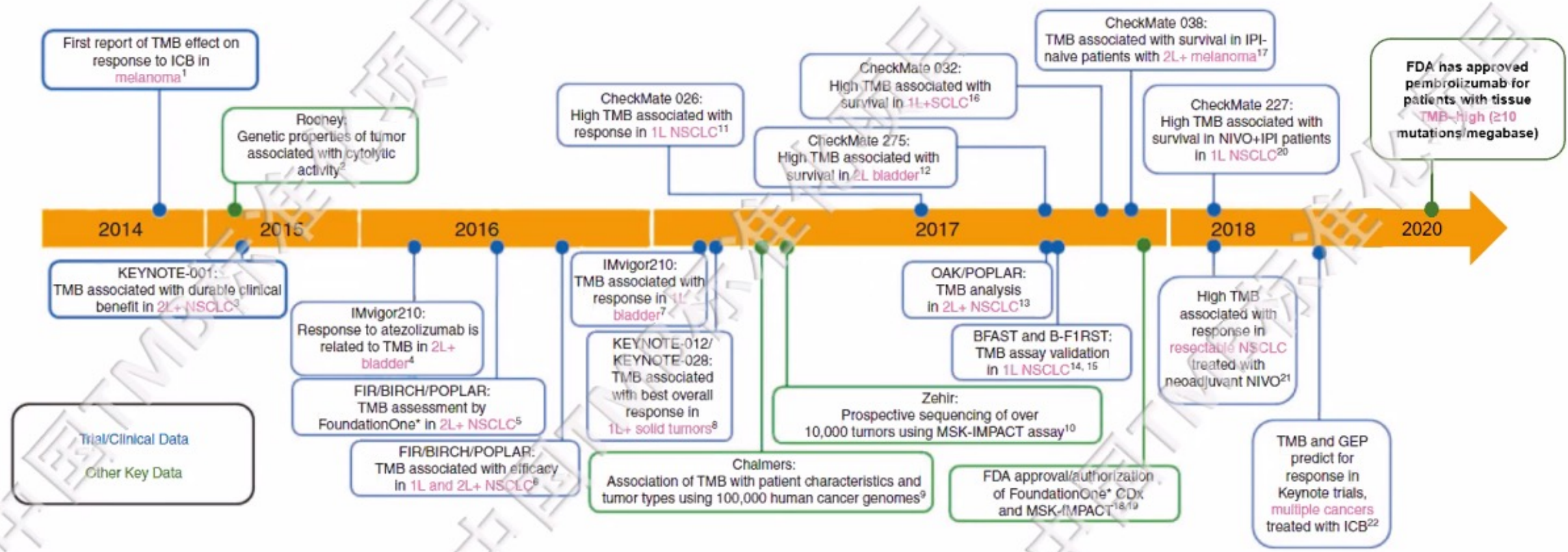
Quality Control and Standardization



中国TMB标准化项目第一阶段总体结果（1A-1B）及1C计划

应建明
国家癌症中心
中国医学科学院
肿瘤医院病理科主任

TMB发展历史

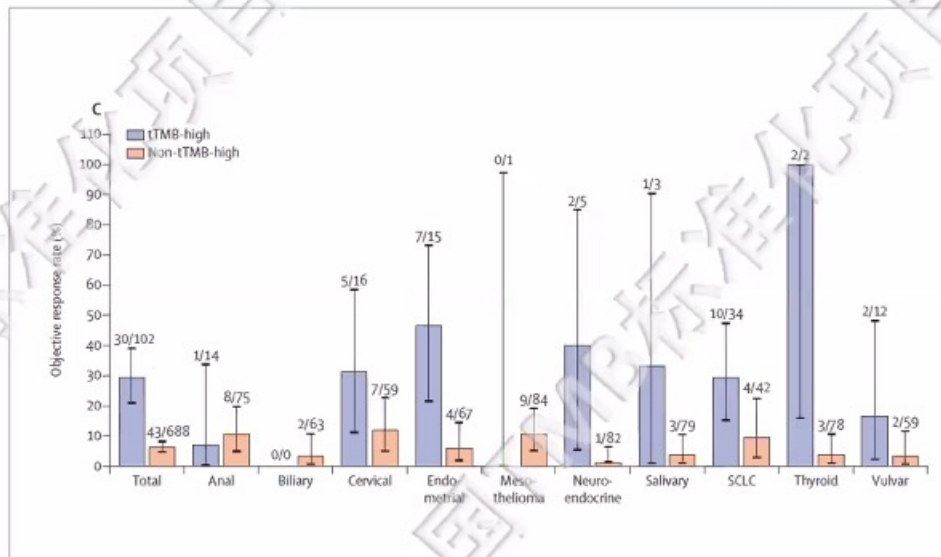


FDA批准TMB作为泛癌种biomarker

TMB-H的ORR高于Non-TMB-H

	tTMB-high (n=102)	tTMB-high (excluding MSI-H; n=81)*	Non-tTMB- high (n=688)
Best response			
Complete response	4 (4%)	3 (4%)	11 (2%)
Partial response	26 (25%)	20 (25%)	32 (5%)
Stable disease	14 (14%)	11 (14%)	227 (33%)
Non-complete response or non-progressive disease†	0	0	3 (<1%)
Progressive disease	48 (47%)	38 (47%)	349 (51%)
Not evaluable‡	1 (1%)	1 (1%)	13 (2%)
Not assessed§	9 (9%)	8 (10%)	53 (8%)
Objective response rate	29% (21-39)	28% (19-40)	6% (5-8)

分癌种的TMB-H和Non-TMB-H的ORR

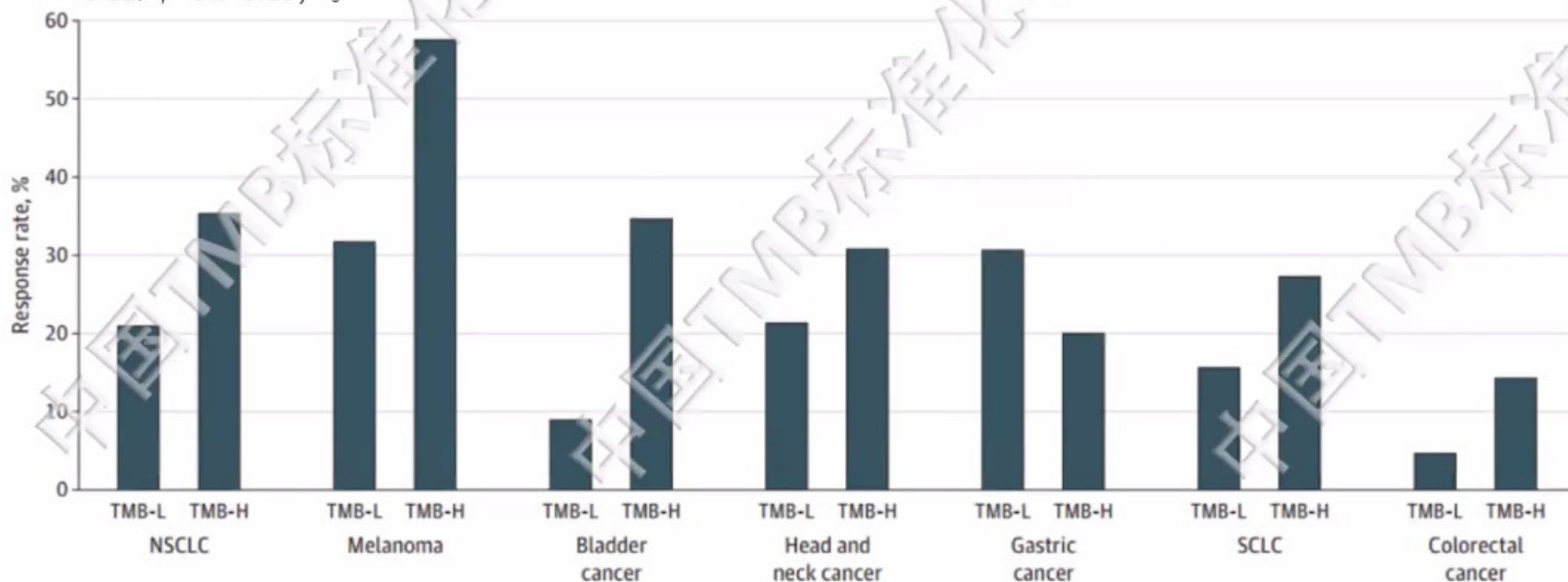


基于Keynote 158临床试验数据结果，FDA批准FoundationOne产品的TMB≥10作为K药治疗晚期癌症患者的biomarker

更多证据说明TMB可有效预测MSS的泛肿瘤患者ICI疗效

研究纳入来自纪念斯隆-凯特琳癌症中心的1678例接受免疫治疗的MSS型晚期实体瘤患者，共纳入16种肿瘤类型，均经MSK-IMPACT或其他FDA批准的NGS平台进行肿瘤组织和配对血液样本的基因检测。

研究分析了之前未经FDA获批免疫治疗适应症的肿瘤ICI治疗的应答率，肿瘤类型包括：肉瘤、结直肠癌、胰腺癌、间皮瘤、卵巢癌、肝内胆管细胞癌及原发灶不明的肿瘤，当10 mut/Mb 作为TMB 通用阈值时，TMB-H肿瘤的整体应答率较TMB-L者高（27% vs 11%，OR=3.15）。



JAMA Oncol. 2021;7(5):739-743. doi:10.1001/jamaoncol.2020.7684

TMB作为泛癌肿biomarker面临的问题

肿瘤突变负荷（TMB）检测及临床应用专家共识

Expert consensus on tumor mutational burden testing and its clinical application

中国抗癌协会肿瘤标志专委会遗传性肿瘤标志物协作组，中国抗癌协会肿瘤病理专业委员会分子病理协作组



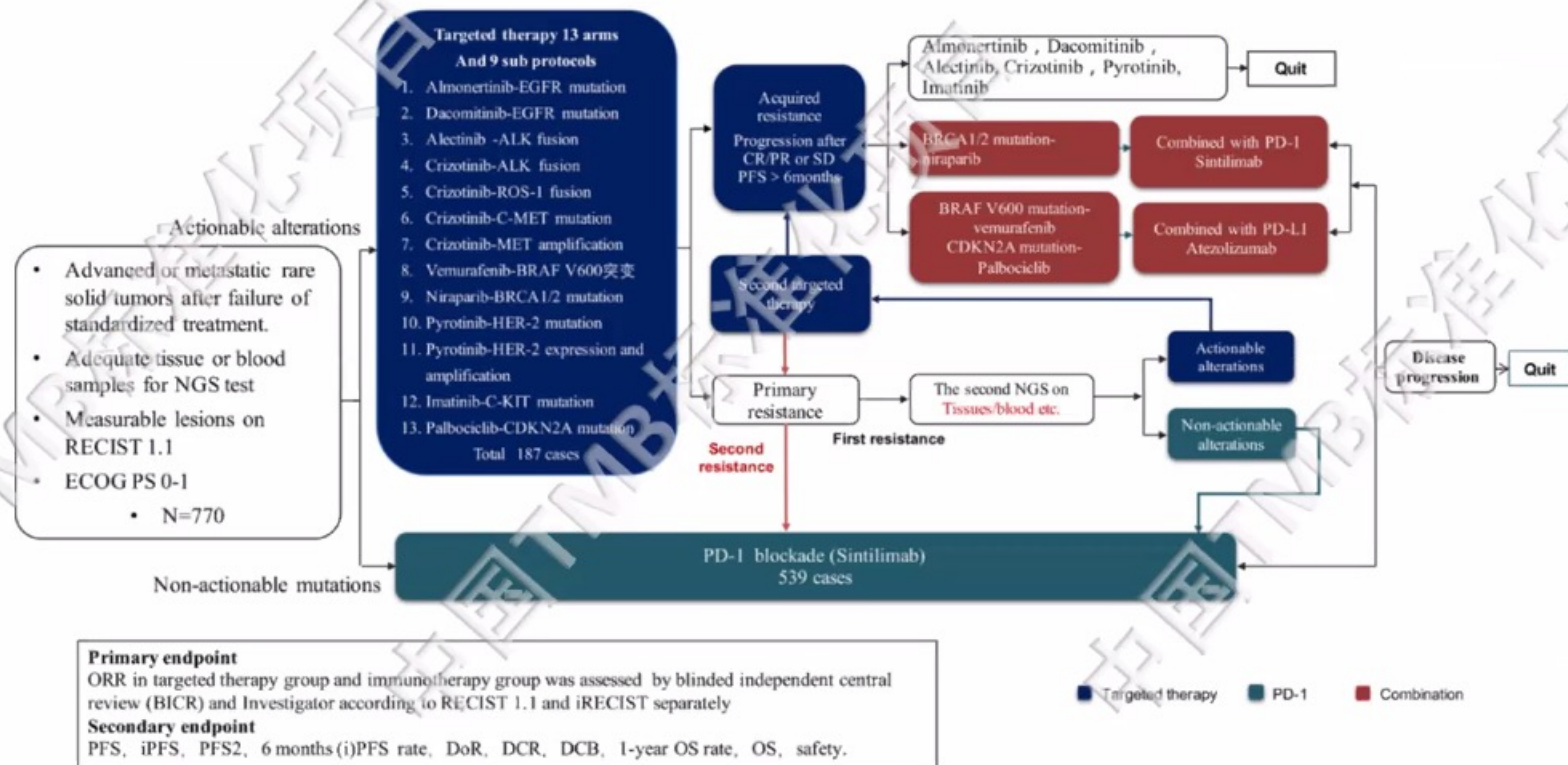
共识:

- 靶向测序Panel覆盖范围原则上不小于1Mb，最低有效测序深度 $\geq 500X$ ；
- TMB的分析纳入 $\geq 5\%$ 的体细胞突变；
- 不同靶向测序Panel的TMB检测体系之间，TMB阈值不能共用；
- TMB的解读建议应用MTB进行临床辅助决策。

问题:

- 产品不同、算法不同，TMB的检测亟需标准化和归一化；
- 依托ICIs疗效随访数据库对基因组比对和突变检出算法开展标准化研究；
- TMB的独立预测价值仍需更多前瞻性研究验证。

PLATFORM临床研究让大量的真实世界样本可以用于TMB项目



中国TMB标准化项目



为更好的满足临床应用需要、契合临床应用场景，规范涉及TMB的多靶点分子标志物检测流程，国家癌症中心中国医学科学院肿瘤医院、中国食品药品检定研究院（中检院）、中国抗癌协会病理专委会共同发起“中国TMB标准化项目（China TMB Harmonization Project）”



临床医生发起的标准临床研究项目

中国TMB标准化项目的核心目标是解决临床实践问题

临床问题

无法评估TMB检测优劣

TMB临床获益阈值不明确

缺乏中国人群TMB临床数据

临床研究

TMB一致性评估与标准化研究

分癌种TMB中国人群基线研究

临床实践

指导ICIs精准用药

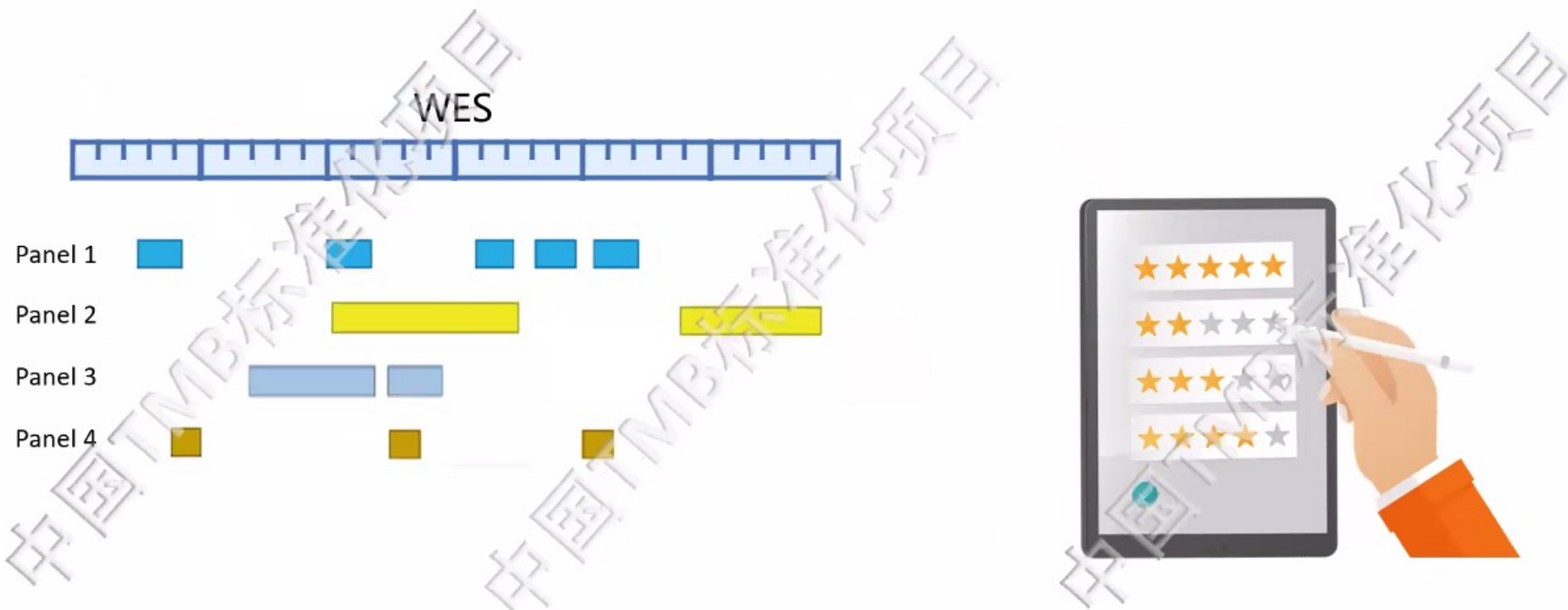
中国TMB标准化项目特色——多维度评估



多biomarker评估，建立分析行业标准，为临床患者提供更好的服务



建立NGS测序panel标准化评估体系



建立以WES为标尺的数据库

建立靶向测序panel标准化评估体系

建立多中心Biomarker真实世界研究科研协作网络



中国TMB标准化项目总览与时间线

项目立项与报名	第一阶段			第二阶段
1. 伦理申请	A: 模拟数据评估	B: 细胞系参考品评估	C: 临床样本评估	D: 分癌种启动 TMB cut-off值构建 临床研究
2. 卫健委备案				
3. 注册临床试验	应用大样本量 WES数据评估参与 单位TMB产品与 WES拟合情况, 统一方法建立各 产品区域校正参 数	通过突变结果统计 分析及湿实验验证 建立标准品答案集, 评估各产品位点检 测准确度、TMB一 致性等指标	评估各单位位点 检出准确度、 TMB一致性和解 读准确度, 验证 统一构建的校正 参数, 建立TMB 标准化评估体系	选择8-10个癌种, 收集带有ICI用药随 访数据样本, 进行 TMB检测及标准化 的评估, 建立药物 疗效cut-off值
4. 与各参与方签订 项目合作协议				
5. 各项准备工作				

项目工作组专家介绍



应建明

国家癌症中心
中国医学科学院
肿瘤医院病理科主任

项目负责人



李宁

国家癌症中心
中国医学科学院肿瘤医院河北医院院长
中国医学科学院肿瘤医院院长助理
临床药物试验中心主任

Platform临床试验发起人



黄杰

中国食品药品检定研究院
体外诊断试剂检定所研究员

TMB参考品评估



周晓燕

复旦大学肿瘤医院
病理科副主任

临床解读评估

项目工作组专家介绍



张绪超

广东省肺癌研究所所长
中心实验室主任

临床解读评估



叶凯

西安交通大学教授
青年千人计划学者

生信分析评估



李程

北京大学生命科学学院
统计科学中心
生物信息中心研究员
国家杰出青年基金获得者

研究设计与统计分析



凌少平

志诺维思
首席执行官
首席科学官

数据分析及平台建设

签订数据安全协议+盲测保证安全性、公平性



参与单位

分别与抗癌登月数据平台签订
数据安全协议;



ClinicalTrials.gov

签署数据安全协议的参与单位
将进入临床研究方案



NCC对参与单位进行盲测编号,
下发邮箱账号;



全球范围内规模最大TMB标准化项目

参与项目的部分医疗机构



科研机构/高校



参与项目的产业机构



参与单位数世界之最

技术评估单位



项目第一阶段时间线



A、模拟数据预评估——提交产品参数



参与单位

提交产品信息

- 专注的癌种;
- Bed文件;
- LOD;
- 有效测序深度;
- 检测突变类型;
- 测序数据量;
- 报告时间;
- 其他指标;

CANCER 抗癌登月
www.0cancer.cn



Cancer Gene Testing Analysis (CGTA)

B、参考品评估

中检院提供参考品 DNA
(500ng的肿瘤参考品和对照
参考品)，同时使用高深度
WES做TMB参考标准



Cell line参考品

WES

共计20个参考品

Panel 1

Panel 2

Panel 3

Panel 4



DNA下发各参与单
位，各参与单位使
用公司产品测序

建库测序



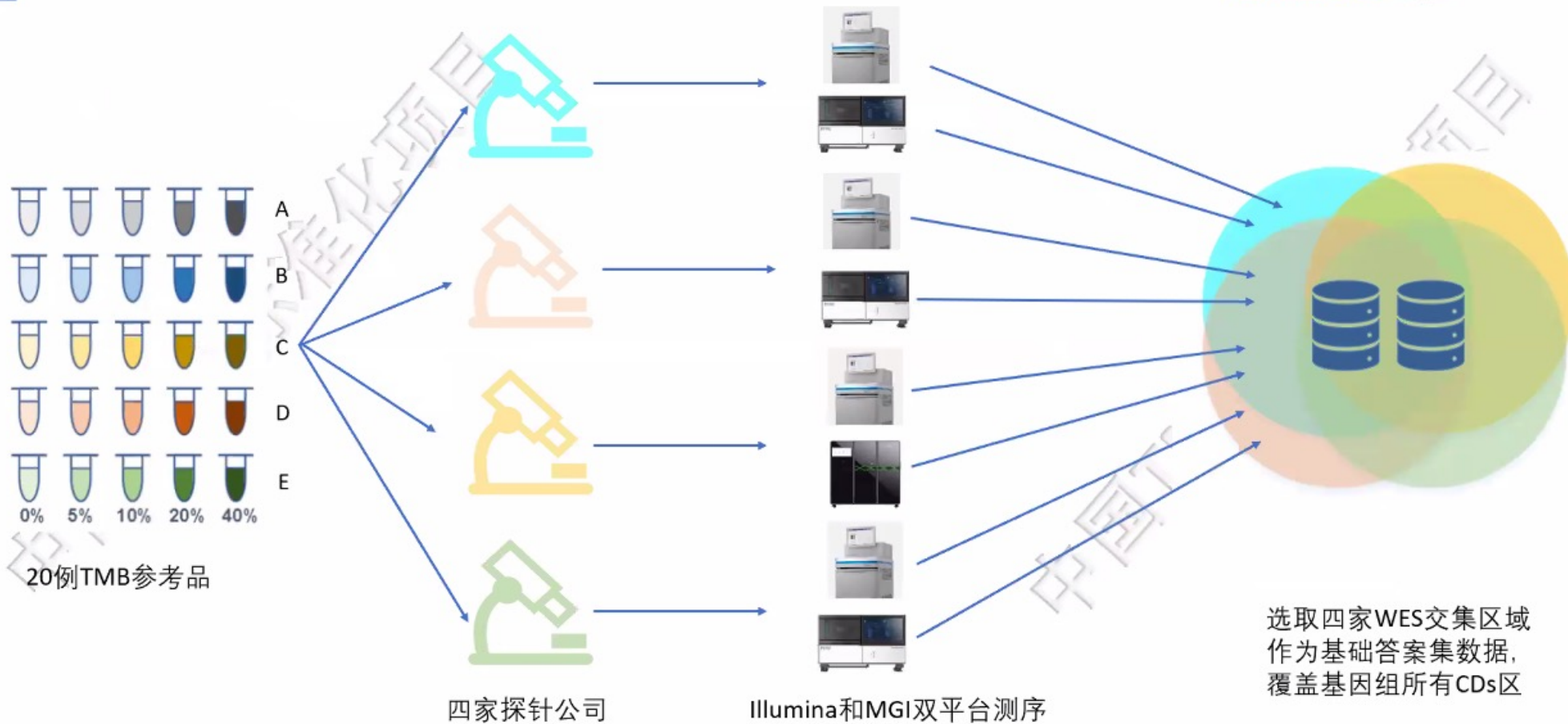
Cancer Gene Testing Analysis (CGTA)



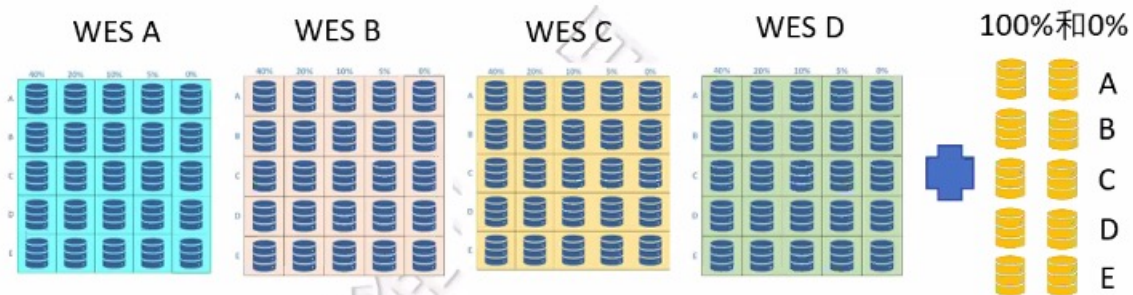
上报测序数据fastq/bam、vcf (14
天上传vcf, 30内上传fastq/bam)



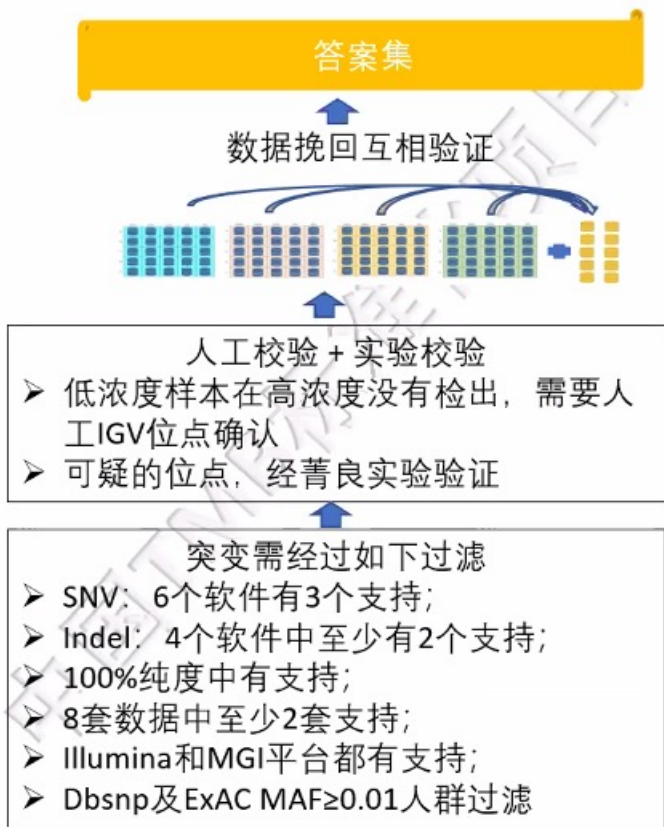
标品采用4家WES探针，双平台进行测序获得数据集



答案集构建流程

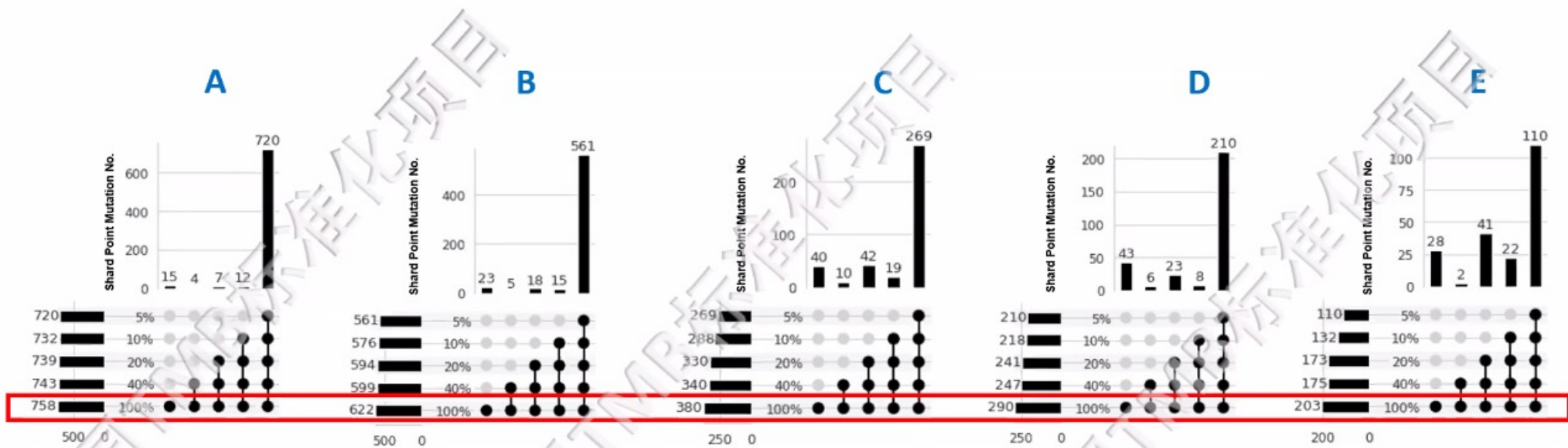


ILMN + MGI 双平台测序数据



答案集经过**专家组共同设计和认证**，根据答案构建流程筛选得到最终每个样本答案的突变位点。

不同浓度的参考品突变统计



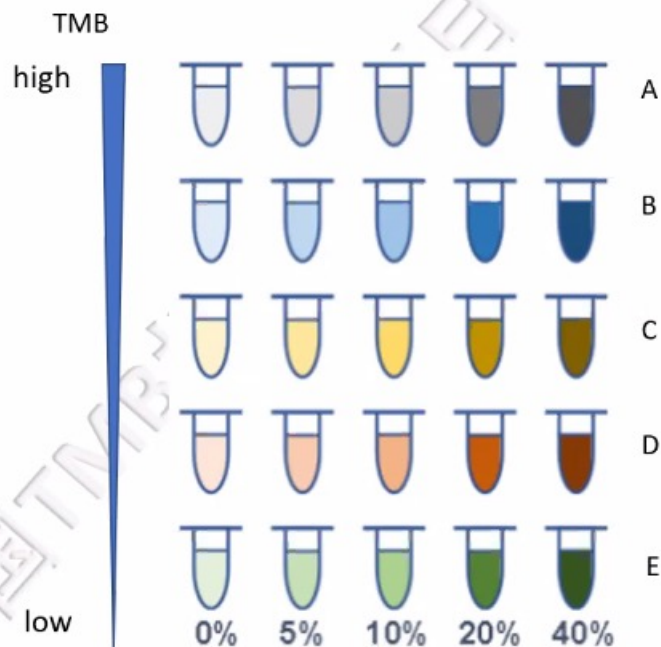
随着肿瘤纯度的下降，肿瘤突变数量逐级降低，体现答案集构建准确性。

从每个参考品来自与一个细胞系的角度出发，使用100%肿瘤比例的答案集结果作为TMB评估的参考值。

(答案集过滤人群SNPs、同义突变，突变位于CDs区域，突变丰度 $\geq 1\%$)

Unpublished data 未经同意，不得使用传播

项目使用不同肿瘤细胞浓度、不同TMB值的参考品



每套TMB标准品包含5组共25支，每组5支。TMB标准品均由肿瘤细胞和配对的正常细胞混合而成，其中肿瘤细胞比例分别为40%、20%、10%、5%和0%。

No.	purity	ID	100%肿瘤比例参考品TMB
1	40%	A40	24.047
2	20%	A20	24.047
3	10%	A10	24.047
4	5%	A05	24.047
5	0%	A0	0
6	40%	B40	19.167
7	20%	B20	19.167
8	10%	B10	19.167
9	5%	B05	19.167
10	0%	B0	0
11	40%	C40	10.370
12	20%	C20	10.370
13	10%	C10	10.370
14	5%	C05	10.370
15	0%	C0	0
16	40%	D40	7.416
17	20%	D20	7.416
18	10%	D10	7.416
19	5%	D05	7.416
20	0%	D0	0
21	40%	E40	4.751
22	20%	E20	4.751
23	10%	E10	4.751
24	5%	E05	4.751
25	0%	E0	0

Unpublished data 未经同意，不得使用传播

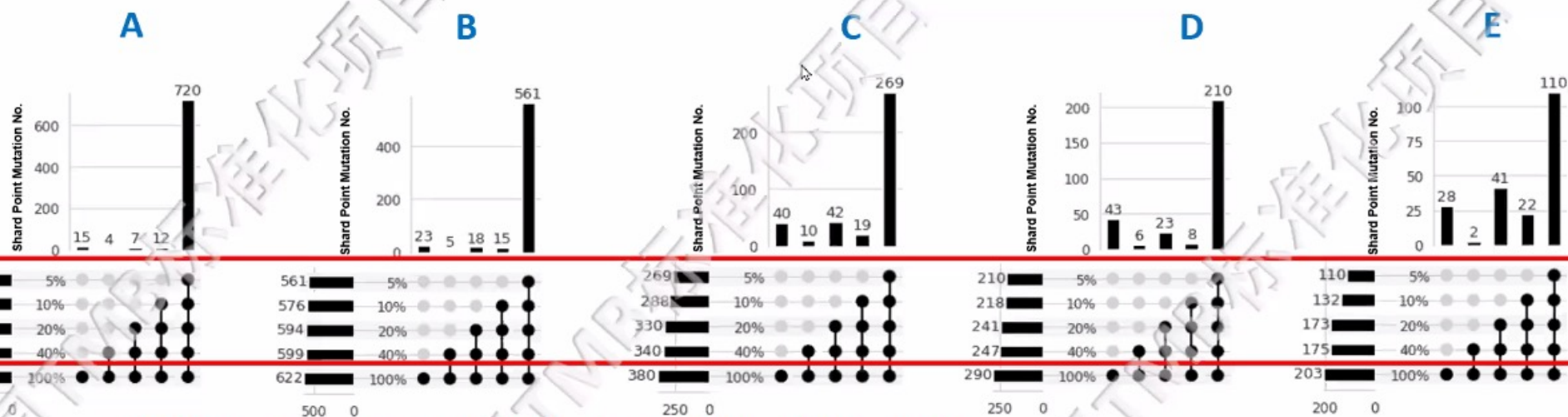


突变位点检出准确度评估

中国TMB标准化项目

中国TMB标准化项目

用于评估各参会单位位点准确率的答案集合 (包含40%, 20%, 10%和5%四种纯度)



所有参与单位提交的40%、20%、10%、5%纯度样本 (ABCDE并集) vcf数据与答案集的对比结果

TP: 真阳, 参与单位与答案集共同检出;

FN: 假阴, 参与单位未检出, 答案集检出;

FP: 假阳, 参与单位检出, 答案集未检出;

PRECI: 精确率, $TP/(TP+FP)$

RECALL: 召回率, $TP/(TP+FN)$

FSCORE: F1值, $2 * PRECI * RECALL / (PRECI + RECALL)$

21家参与单位SNV准确度性能评估

序号	公司编号	声称测序深度	实际测序深度	声称 SNV LOD	实际上SNV的最低VAF	TP	FN	FP	PRECISION	RECALL	F1-SCORE
1	Tmb046	700	3515±591	5.00%	0.10%	201	2	1	99.50%	99.00%	99.30%
2	Tmb015	1000	4403±408	5.00%	0.40%	220	4	5	97.80%	98.20%	98.00%
3	Tmb042	500	2512±189	2.00%	0.90%	131	3	0	100.00%	97.80%	98.90%
4	Tmb009	1000	2575±409	1.00%	0.10%	152	4	3	98.10%	97.40%	97.70%
5	Tmb040	1000	2005±325	1.00%	0.40%	152	4	3	98.10%	97.40%	97.70%
6	Tmb043	1200	4972±651	3.00%	5.00%	291	8	1	99.70%	97.30%	98.50%
7	Tmb020	500	1734±300	2.00%	0.70%	170	5	0	100.00%	97.10%	98.60%
8	Tmb025	250	1802±94	2.00%	1.30%	160	7	7	95.80%	95.80%	95.80%
9	Tmb014	1000	1838±160	2.00%	2.00%	105	5	0	100.00%	95.50%	97.70%
10	Tmb014	500	2599±430	2.00%	0.40%	166	8	0	100.00%	95.40%	97.60%
11	Tmb022	500	1534±141	2.00%	1.80%	198	10	2	99.00%	95.20%	97.10%
12	Tmbxxx	500	3035±317	0.40%	0.70%	208	14	7	96.70%	93.70%	95.20%
13	Tmbxxx	500	2256±486	2.00%	2.00%	133	13	1	99.30%	91.10%	95.00%
14	Tmbxxx	500	2434±332	2.00%	2.00%	134	14	0	100.00%	90.50%	95.00%
15	Tmbxxx	500	1020±261	1.00%	0.20%	215	27	4	98.20%	88.80%	93.30%
16	Tmbxxx	1600	1396±148	1.00%	0.70%	207	34	1	99.50%	85.90%	92.20%
17	Tmbxxx	700	2208±406	5.00%	0.40%	171	34	0	100.00%	83.40%	91.00%
18	Tmbxxx	450	1070±315	1.00%	0.50%	316	69	4	92.80%	82.10%	89.60%
19	Tmbxxx	500	1461±166	0.50%	1.00%	191	60	0	100.00%	76.10%	86.40%
20	Tmbxxx	500	968±107	5.00%	5.00%	21	53	21	50.00%	28.40%	36.20%
21	Tmbxxx	2500	3530±375	1.00%	1.00%	67	182	1	98.50%	26.90%	42.30%

所有样本SNV，按照参与单位上报的LOD进行位点准确性评估（答案位点VAF最低为1%）

Unpublished data 未经同意，不得使用传播

11家SNV-LOD达标单位的SNV准确度性能展示

序号	公司编号	声称 测序深度	实际 测序深度	声称 SNV-LOD	实际上报SNV 的最低VAF	TP	FN	FP	PRECISION	RECALL	F1-SCORE
1	tmb040	1000	2005±325	1.00%	0.10%	152	4	3	98.10%	97.40%	97.70%
2	tmb009	1000	2575±409	1.00%	0.40%	152	4	3	98.10%	97.40%	97.70%
3	tmb025	250	1802±94	2.00%	1.30%	160	7	7	95.80%	95.80%	95.80%
4	tmb022	500	1534±141	2.00%	1.80%	198	10	2	99.00%	95.20%	97.10%
5	tmb014	1000	1838±160	2.00%	2.00%	105	5	0	100.00%	95.50%	97.70%
6	tmb020	500	1734±300	2.00%	0.70%	170	5	0	100.00%	97.10%	98.60%
7	tmb024	500	2599±430	2.00%	0.40%	166	8	0	100.00%	95.40%	97.60%
8	tmb042	500	2512±189	2.00%	0.90%	131	3	0	100.00%	97.80%	98.90%
9	tmb043	1200	4972±651	3.00%	5.00%	291	8	1	99.70%	97.30%	98.50%
10	tmb015	1000	4403±408	5.00%	0.40%	220	4	5	97.80%	98.20%	98.00%
11	tmb046	700	3515±591	5.00%	0.10%	201	2	1	99.50%	99.00%	99.30%

Recall≥95%的参与单位，按照1) 声称SNV-LOD升序；2) F1-score 降序排序

其他参与单位SNV准确度评估及GVC性能展示



参与单位生信流程

GVC生信流程性能

公司编号	声称测序深度	实际测序深度	声称 SNV LOD	实际上SNV的最低频率	TP	FN	FP	PRECI	RECALL	FSCORE	TP	FN	FP	PRECI	RECALL	F1-SCORE	F1-SCORE 差值
tmbxxx	2500	3530±375	1.00%	1.00%	67	182	1	98.50%	26.90%	42.30%	237	12	5	97.90%	95.20%	96.50%	54.27%
tmbxxx	500	2256±486	2.00%	2.00%	133	13	1	99.30%	91.10%	95.00%	144	3	1	99.30%	98.00%	98.60%	3.63%
tmbxxx	500	2434±332	2.00%	2.00%	134	14	0	100.00%	90.50%	95.00%	140	7	0	100.00%	95.20%	97.60%	2.53%
tmbxxx	700	2208±406	5.00%	0.40%	171	34	0	100.00%	83.40%	91.00%	207	2	0	100.00%	99.00%	99.50%	8.56%
tmbxxx	500	968±107	5.00%	5.00%	21	53	21	50.00%	28.40%	36.20%	74	2	0	100.00%	97.40%	98.70%	62.46%
tmbxxx	500	3035±317	0.40%	0.70%	208	14	7	96.70%	93.70%	95.20%	209	13	8	96.30%	94.10%	95.20%	0.02%
tmbxxx	500	1461±166	0.50%	1.00%	191	60	0	100.00%	76.10%	86.40%	232	19	12	95.10%	92.40%	93.70%	7.31%
tmbxxx	500	1020±261	1.00%	0.20%	215	27	4	98.20%	88.80%	93.30%	218	25	7	96.90%	89.70%	93.20%	-0.11%
tmbxxx	1600	1396±148	1.00%	0.70%	207	34	1	99.50%	85.90%	92.20%	227	17	9	96.20%	93.00%	94.60%	2.38%
tmbxxx	450	1070±315	1.00%	0.50%	316	69	4	98.80%	82.10%	89.60%	334	52	10	97.10%	86.50%	91.50%	1.86%

Unpublished data 未经同意，不得使用传播

12家INDEL-LOD达标单位INDEL准确度评估与GVC性能展示



参与单位生信流程

GVC生信流程性能

序号	公司编号	声称测序深度	实际测序深度	声称 INDEL-LOD	实际上报INDEL的最低VAF	TP	FN	FP	PRECI	RECALL	F1-SCORE	TP	FN	FP	PRECISION	RECALL	FSCORE	F1score 差值
1	tmb027	500	1461±166	0.50%	1.00%	15	0	1	93.75%	100.00%	96.77%	15	0	1	93.75%	100.00%	96.77%	0.00%
2	tmb040	1000	2005±325	1.00%	0.40%	23	0	1	95.83%	100.00%	97.87%	23	0	0	100.00%	100.00%	100.00%	2.13%
3	Tmb009	1000	2575±409	1.00%	0.10%	23	0	1	95.83%	100.00%	97.87%	24	0	0	100.00%	100.00%	100.00%	2.13%
4	Tmb017	1600	1396±148	1.00%	0.70%	11	0	1	91.67%	100.00%	95.65%	12	0	0	100.00%	100.00%	100.00%	4.35%
5	Tmb035	450	1070±315	1.00%	0.50%	11	0	2	84.62%	100.00%	91.67%	12	0	0	100.00%	100.00%	100.00%	8.33%
6	Tmb044	500	1020±261	1.00%	0.20%	11	0	3	78.57%	100.00%	88.00%	12	0	0	100.00%	100.00%	100.00%	12.00%
7	Tmb020	500	1734±300	2.00%	0.70%	23	0	0	100.00%	100.00%	100.00%	24	0	0	100.00%	100.00%	100.00%	0.00%
8	Tmb024	500	2599±430	2.00%	0.40%	11	0	0	100.00%	100.00%	100.00%	11	0	0	100.00%	100.00%	100.00%	0.00%
9	Tmb046	700	3515±591	5.00%	0.10%	16	0	2	88.89%	100.00%	94.12%	24	0	0	100.00%	100.00%	100.00%	5.88%
10	Tmb022	500	1534±141	5.00%	1.80%	9	0	4	69.23%	100.00%	81.82%	9	0	0	100.00%	100.00%	100.00%	18.18%
11	Tmb042	500	2512±189	5.00%	0.90%	9	0	4	69.23%	100.00%	81.82%	9	0	0	100.00%	100.00%	100.00%	18.18%
12	Tmb015	1000	4403±408	5.00%	0.40%	13	0	23	36.11%	100.00%	53.06%	11	0	0	100.00%	100.00%	100.00%	46.94%

Recall≥95%的参与单位，按照1) 声称SNV LOD升序; 2) F1 score降序排序 3) 序号不代表排名

所有样本Indel，按照参与单位上报的LOD进行位点准确性评估（答案位点频率最低为1%）

Unpublished data 未经同意，不得使用传播

其他参与单位INDEL准确度评估与GVC性能展示

参与单位生信流程

GVC生信流程

公司编号	声称测序深度	实际测序深度	声称 INDEL LOD	实际上报 INDEL 的最低 VAF	TP	FN	FP	PRECI	RECALL	F1 SCORE	TP	FN	FP	PRECI	RECALL	F1 SCORE	F1score 差值
tmbxxx	500	3035±317	0.40%	0.70%	6	9	0	100.00%	40.00%	57.14%	15	0	4	78.95%	100.00%	88.24%	31.10%
tmbxxx	2500	3530±375	1.00%	1.00%	5	6	0	100.00%	45.45%	62.50%	12	0	0	100.00%	100.00%	100.00%	37.50%
tmbxxx	1000	1838±160	2.00%	2.00%	10	1	0	100.00%	90.91%	95.24%	11	0	0	100.00%	100.00%	100.00%	4.76%
tmbxxx	250	1802±94	2.00%	1.30%	0	7	6	0.00%	0.00%	0.00%	8	0	0	100.00%	100.00%	100.00%	100.00%
tmbxxx	1200	4972±651	3.00%	5.00%	4	26	23	14.81%	13.33%	14.04%	29	1	0	100.00%	96.67%	98.31%	84.27%
tmbxxx	500	2256±486	3.00%	2.00%	2	8	0	100.00%	20.00%	33.33%	10	0	0	100.00%	100.00%	100.00%	66.67%
tmbxxx	500	2434±332	3.00%	2.00%	9	1	0	100.00%	90.00%	94.74%	10	0	0	100.00%	100.00%	100.00%	5.26%
tmbxxx	700	2208±406	5.00%	0.40%	14	2	3	82.35%	87.50%	84.85%	24	0	0	100.00%	100.00%	100.00%	15.15%
tmbxxx	500	968±107	8.00%	5.00%	0	0	0	0.00%	0.00%	0.00%	12	0	0	100.00%	100.00%	100.00%	100.00%

Recall≥95%的参与单位，按照1) 声称SNV LOD升序；2) F1 score降序排序

Unpublished data 未经同意，不得使用传播

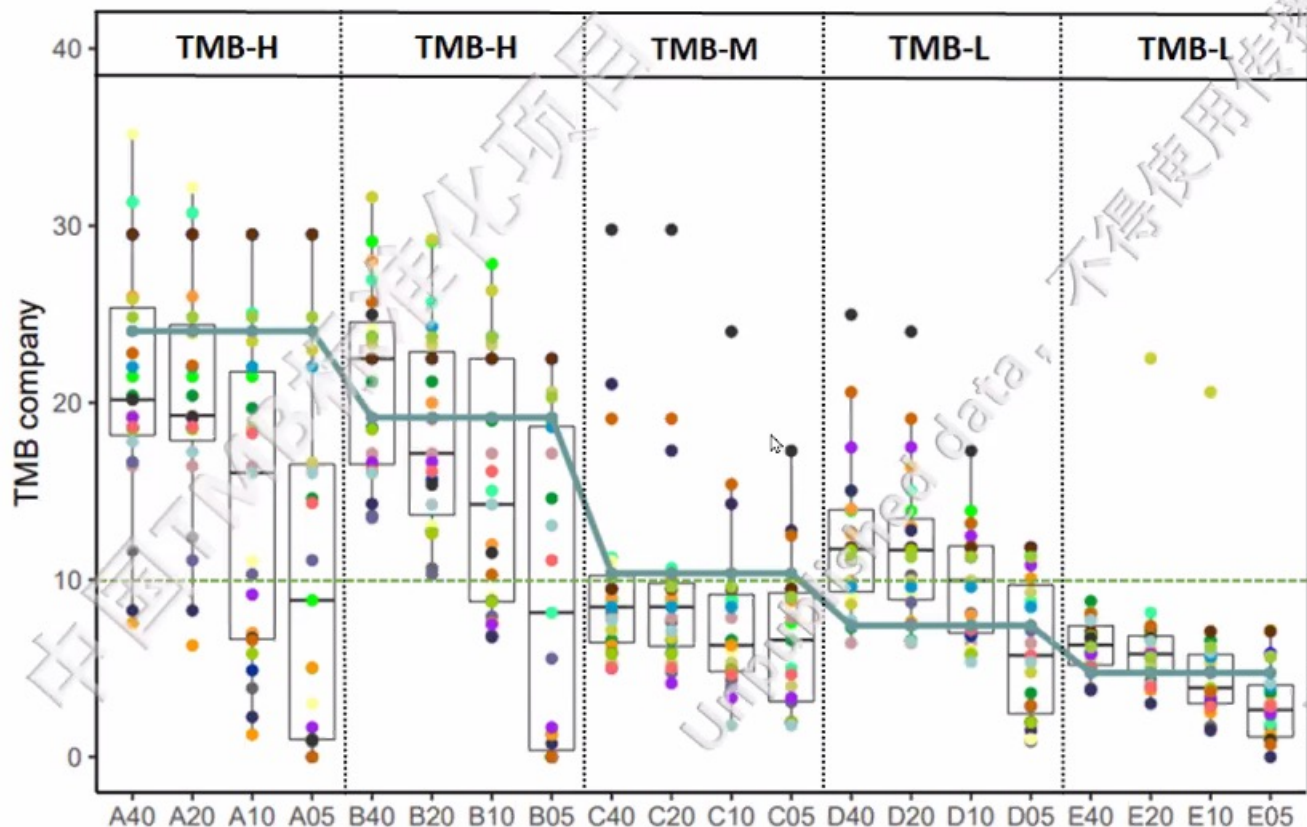
TMB的计算方法

序号	阈值判断方法	SNPs数据库过滤	突变类型	计算TMB VAF	突变检出VAF	最低纯度	是否考虑纯度
tmbXXX	10	COSIMC; 1000人群; dbSNP; ExAC	非同义突变	1%	1%;1%	5%	否
tmbXXX	自建	COSMIC	非同义突变	5%	5%;5%	20%	否
tmbXXX	TCGA	COSIMC; 1000人群; dbSNP; ExAC	非同义突变	1%	1%;1%	20%	否
tmbXXX	自建	1000G和ExAc	仅去除驱动突变	5%	2%;2%	20%	否
tmbXXX	10	否	非同义突变	5%	2%;5%	20%	否
tmbXXX	10	ExAC	非同义突变	1%	2%;2%	20%	否
tmbXXX	自建	dbSNP144、esp6500、1000g、exac、GnomAD	含同义突变	5%	2%;2%	20%	否
tmbXXX	自建	ExAC;gnomAD	含同义突变	1%	1%;1%	50%	否
tmbXXX	自建	NA	非同义突变	NA	0.5%;0.5%	20%	NA
tmbXXX	10	dbSNP; ExAC	含同义突变	5%	5%;8%	20%	是
tmbXXX	10	NA	非同义突变	0.2%	NA	NA	NA
tmbXXX	10	COSMIC;千人;NHLBI;gnomAD;内部	非同义突变	5%;2.5%;1.25%;1%	5%;5%	20%	是
tmbXXX	自建	驱动基因; ExAC	非同义突变	5%	2%;3%	20%	否
tmbXXX	10	COSIMC; 1000人群; dbSNP; ExAC	非同义突变	1%	1%;1%	5%	否
tmbXXX	自建	驱动基因; ExAC	非同义突变	5%	2%;3%	20%	否
tmbXXX	自建	否	非同义突变	5%	2%;5%	20%	否
tmbXXX	自建	千人数据-Exac人群免疫-驱动数据库	含同义突变	1.5%	3%;3%	20%	NA
tmbXXX	16	NA	非同义突变	5%	1%;1%	20%	NA
tmbXXX	TCGA	1000Genome;ExAC;ESP6500	非同义突变	1%	0.4%;0.4%	10%	否
tmbXXX	10	COSMIC+内部数据库	非同义突变	5%;2.5%;1.25%;1%	5%;5%	20%	是

各参与单位计算TMB的方法及阈值较为多样;

统一分析流程: 去除dbSNP+ExAC上MAF \geq 0.01的SNPs; 去除同义突变; 突变丰度5%以上cds区域突变

各参与单位上报TMB值及100%浓度TMB答案值分布图



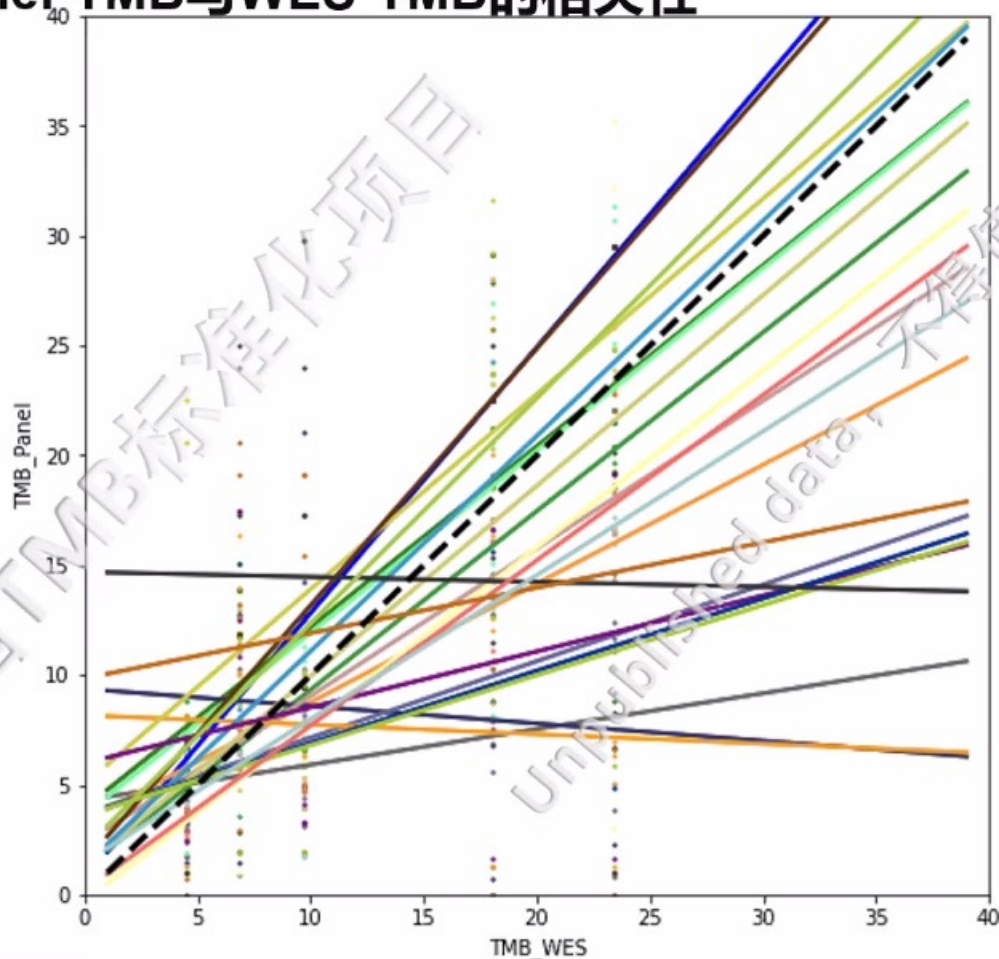
Company No.

- tmb009
- tmb014
- tmb015
- tmb017
- tmb019
- tmb020
- tmb022
- tmb024
- tmb025
- tmb026
- tmb027
- tmb033
- tmb035
- tmb036
- tmb037
- tmb039
- tmb040
- tmb041
- tmb042
- tmb043
- tmb044
- tmb045
- tmb046

绿色折线为100%肿瘤细胞浓度下TMB参考值 (计算TMB的VAF \geq 5%) :

参考品编号	100%肿瘤细胞浓度下TMB参考值
A	24.047
B	19.167
C	10.370
D	7.416
E	4.751

Panel-TMB与WES-TMB的相关性

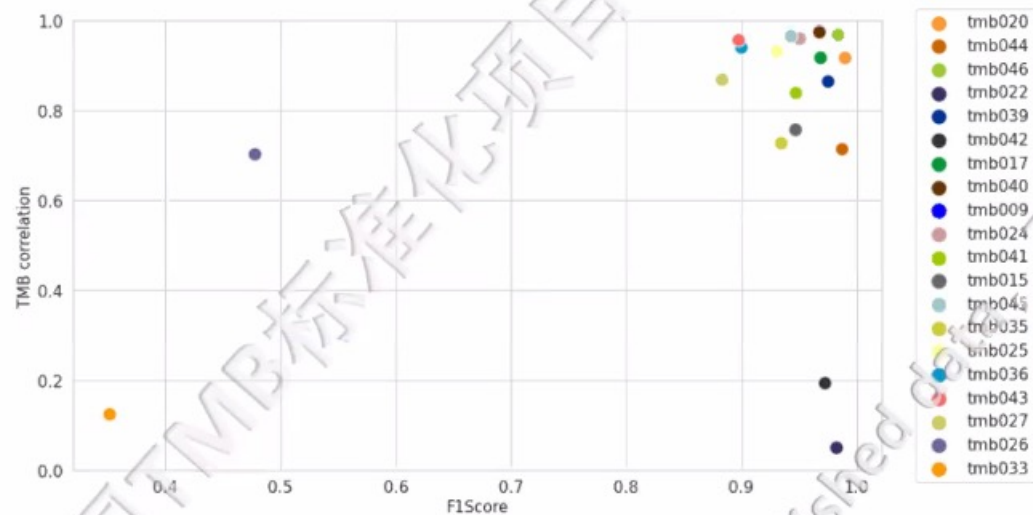


tmb009	$y=1.209x+0.738$	$r=0.977$	$p=0.000$
tmb040	$y=1.170x+1.477$	$r=0.976$	$p=0.000$
tmb046	$y=1.025x+2.109$	$r=0.963$	$p=0.000$
tmb024	$y=0.672x+2.327$	$r=0.954$	$p=0.000$
tmb036	$y=0.980x+1.300$	$r=0.942$	$p=0.000$
tmb043	$y=0.752x+0.180$	$r=0.935$	$p=0.000$
	$y=0.813x+1.202$	$r=0.896$	$p=0.000$
	$y=0.655x+1.445$	$r=0.895$	$p=0.000$
	$y=0.868x+1.269$	$r=0.846$	$p=0.000$
	$y=0.824x+3.942$	$r=0.715$	$p=0.000$
	$y=0.889x+5.020$	$r=0.709$	$p=0.000$
	$y=0.830x+3.618$	$r=0.667$	$p=0.001$
	$y=0.348x+3.673$	$r=0.645$	$p=0.002$
	$y=0.806x+0.310$	$r=0.595$	$p=0.006$
	$y=0.541x+3.351$	$r=0.455$	$p=0.044$
	$y=0.319x+3.613$	$r=0.395$	$p=0.085$
	$y=0.328x+3.631$	$r=0.390$	$p=0.089$
	$y=0.255x+5.982$	$r=0.288$	$p=0.218$
	$y=0.163x+4.291$	$r=0.280$	$p=0.233$
	$y=0.206x+9.843$	$r=0.180$	$p=0.447$
	$y=-0.023x+14.692$	$r=-0.016$	$p=0.945$
	$y=-0.043x+8.168$	$r=-0.068$	$p=0.776$
	$y=-0.079x+9.370$	$r=-0.088$	$p=0.711$

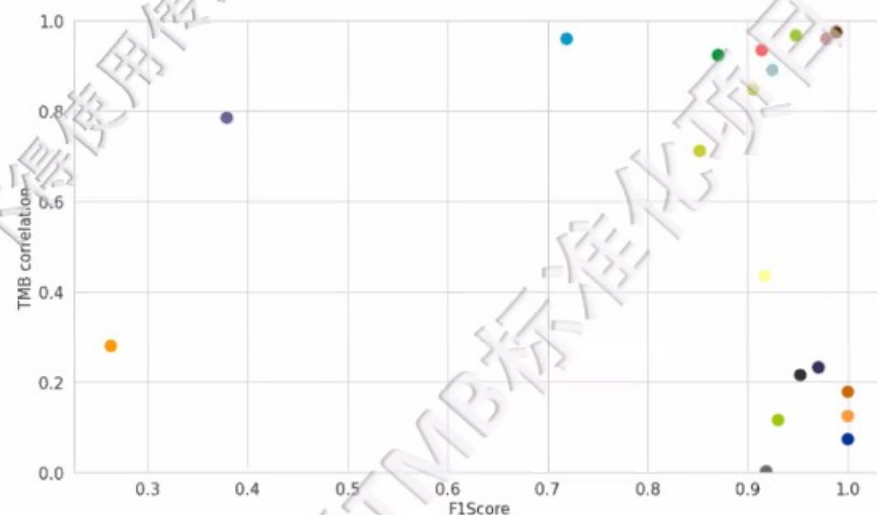
横坐标为100%肿瘤细胞比例下WES-TMB (VAF≥5%)
纵坐标为Panel-TMB

Unpublished data 未经同意, 不得使用传播

综合分析突变检测准确度与TMB数值结果

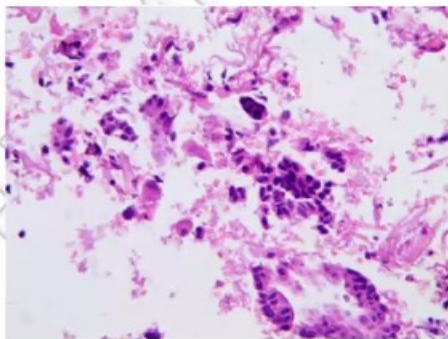
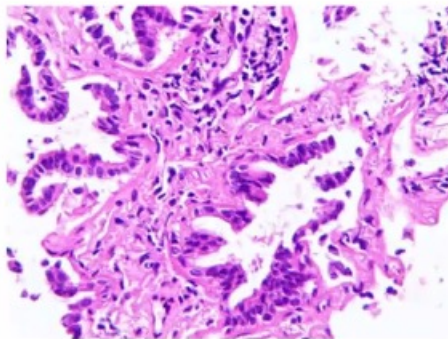


肿瘤纯度40%和20%的标品
(F1-Score根据各家实际计算TMB的VAF来统计)



肿瘤纯度10%和5%的标品
(F1-Score根据各家实际计算TMB的VAF来统计)

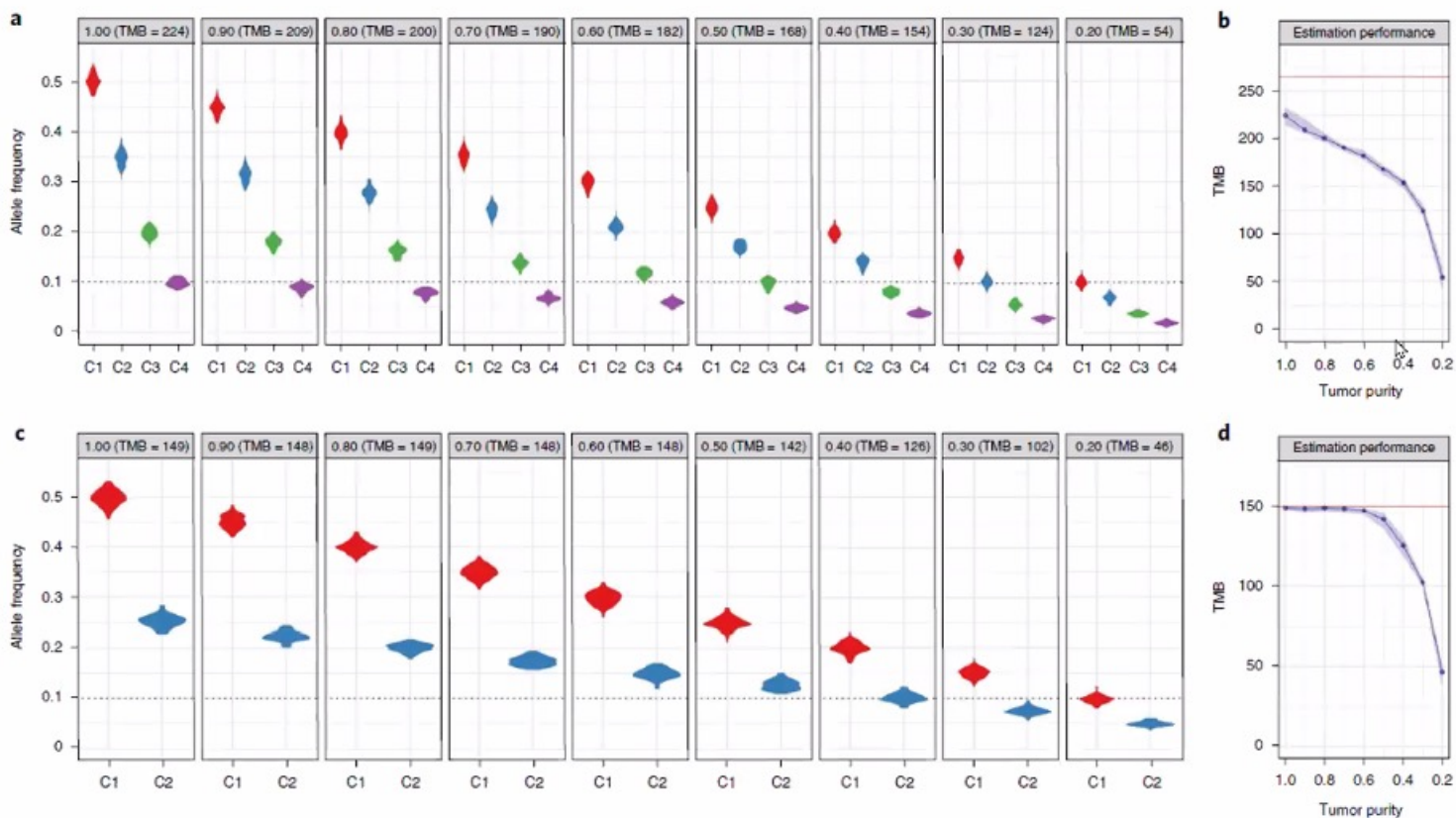
横坐标为位点检出准确度评估指标F1 score，纵坐标为Panel-TMB与WES-TMB相关性的评估指标TMB correlation
当肿瘤纯度 $\geq 20\%$ ，绝大部分参与单位都有很好的表现，二者成正比



肿瘤纯度对样本TMB计算的影响评估

中国TMB标准化项目

肿瘤纯度对TMB计算的影响不容忽视

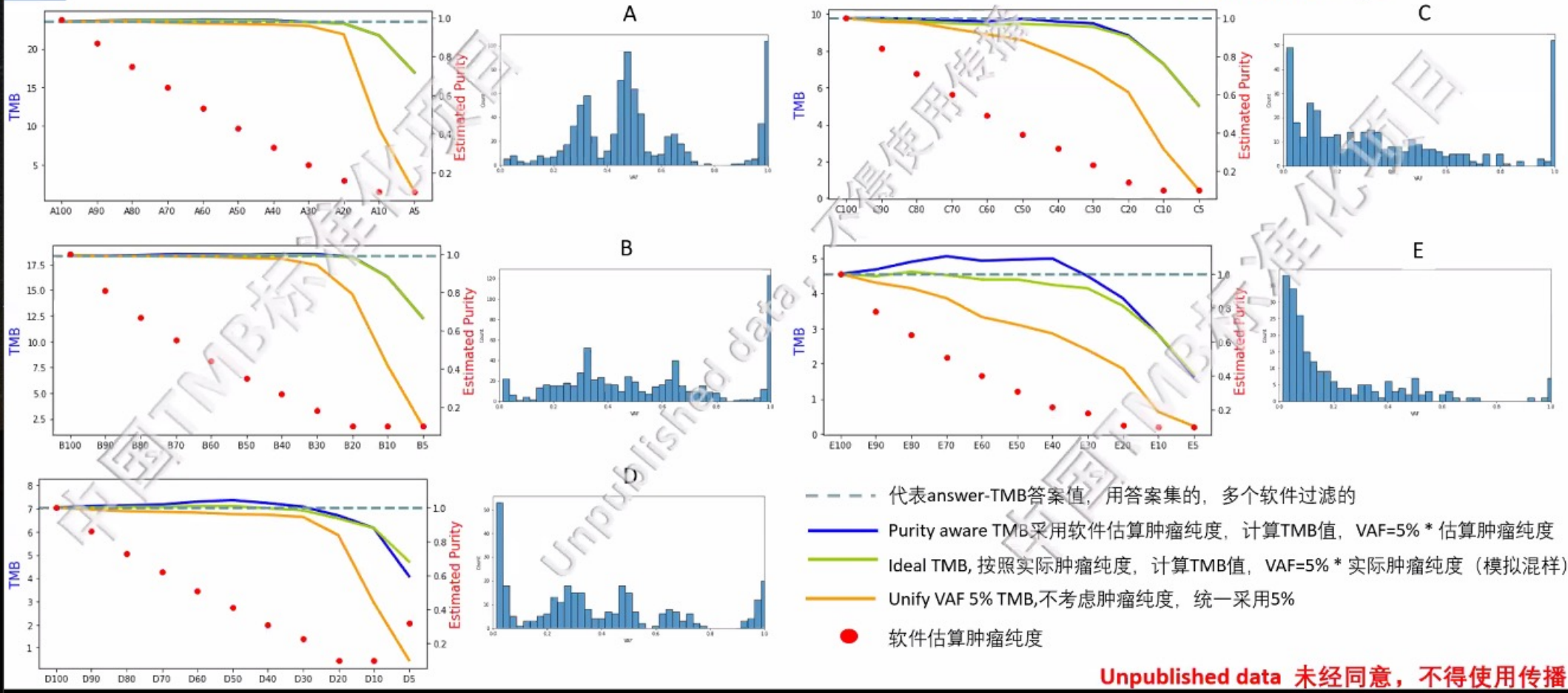


样本中低肿瘤纯度可导致TMB检测敏感度降低。

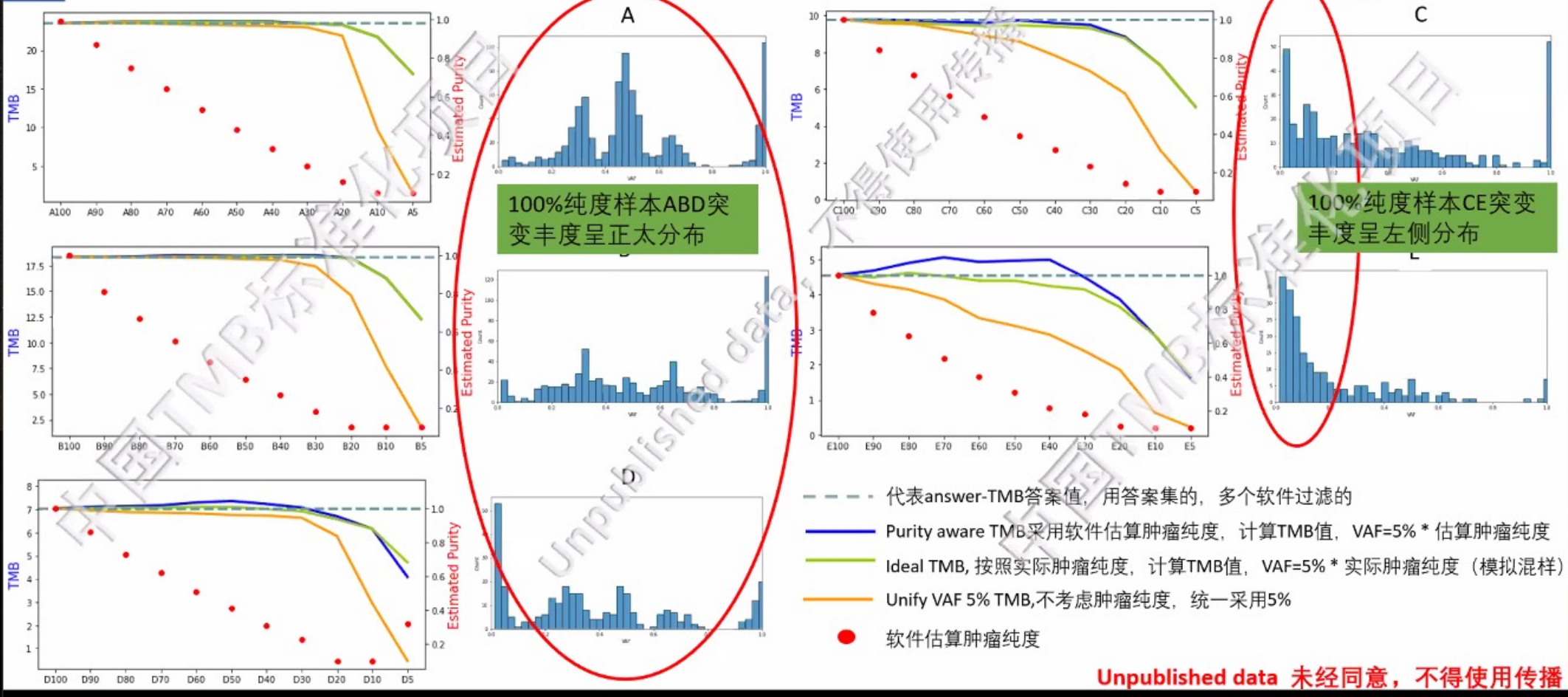
思考：通过降低TMB计算VAF减少低肿瘤纯度对TMB的影响

- 目前采用统一的5%的VAF来计算TMB，但受肿瘤纯度的影响，低肿瘤纯度样本位点突变丰度普遍降低，从而引起样本TMB的降低，使患者失去用药机会；
- 各参与单位VAF在1%以上已经有比较高的位点检出准确率，那是否可以通过降低LOD的方式，纳入低肿瘤纯度样本中低于5%的突变位点，是否可以还原TMB，增加患者的用药机会？

肿瘤纯度的降低使得TMB值失真严重

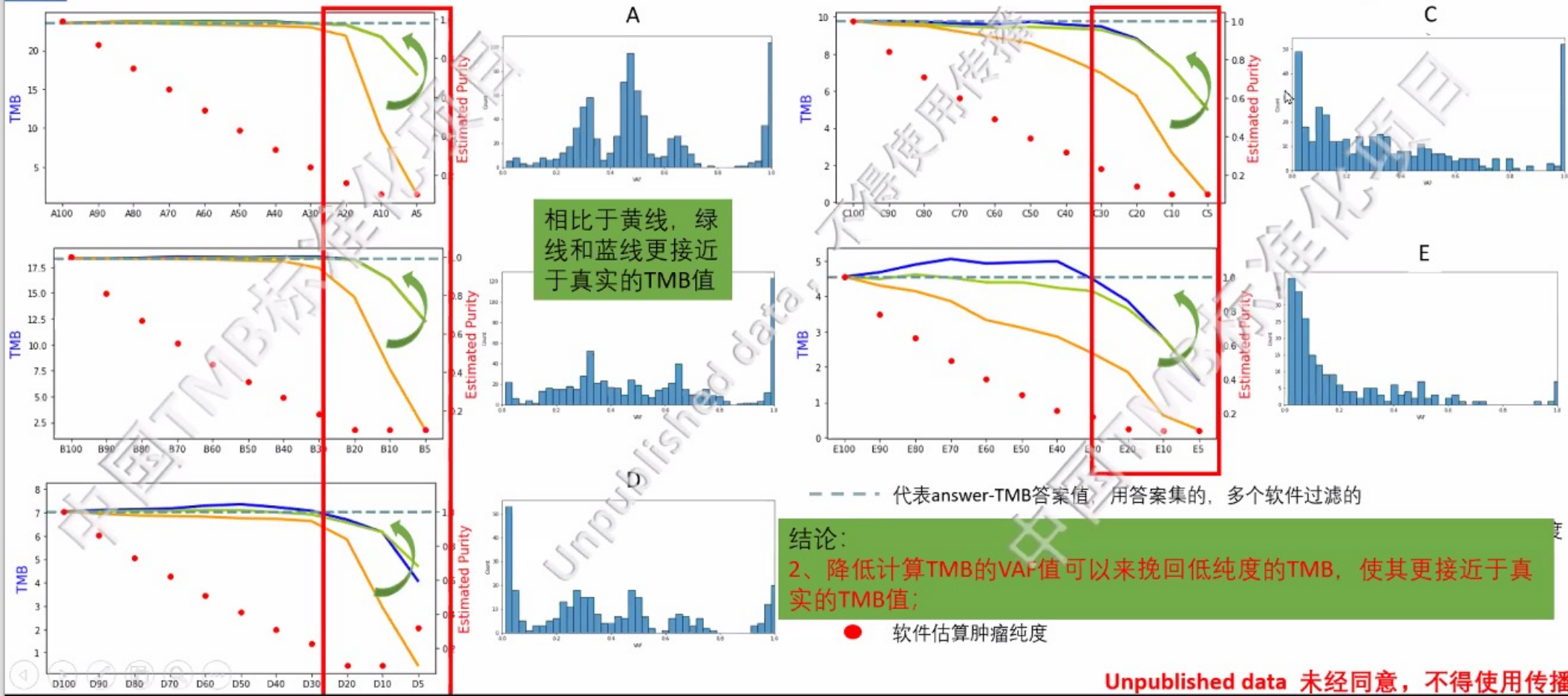


肿瘤纯度的降低使得TMB值失真严重

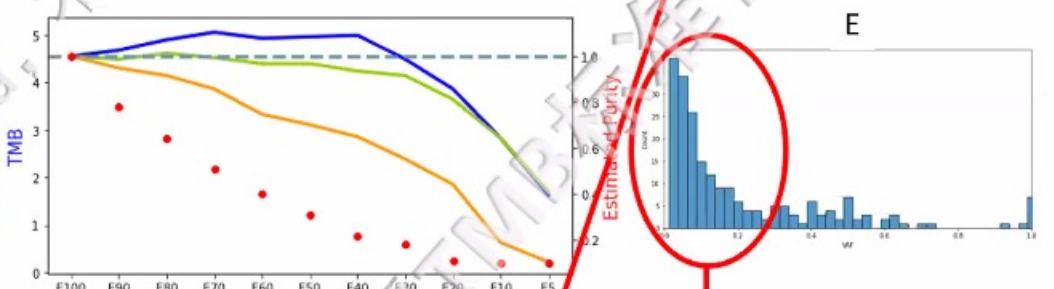
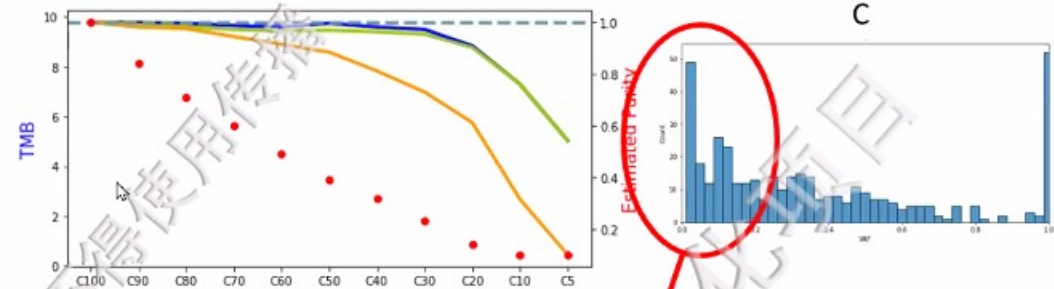
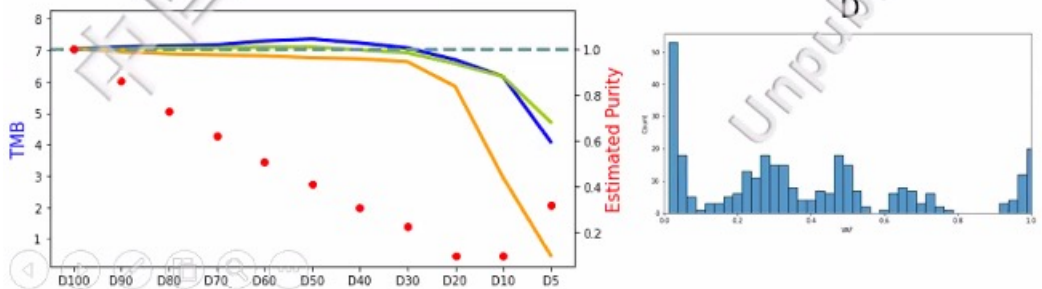
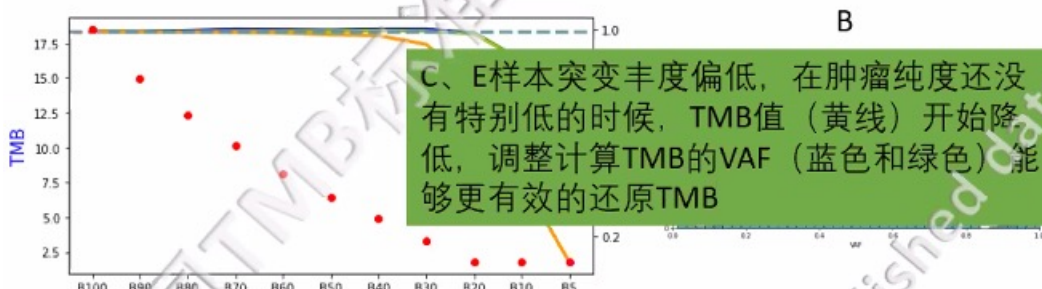
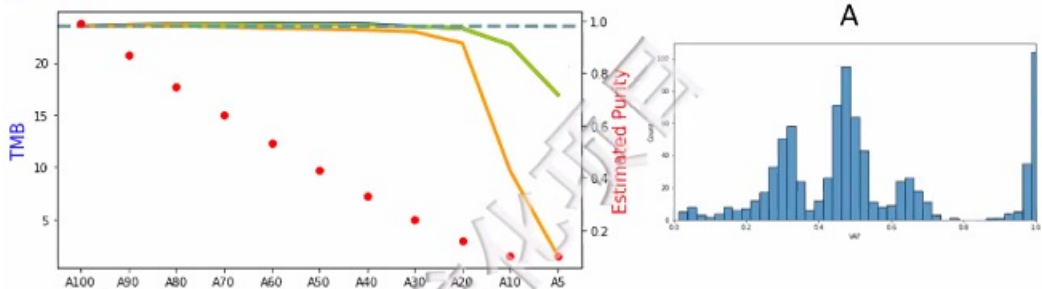


Unpublished data 未经同意, 不得使用传播

降低计算TMB的VAF值可减低肿瘤纯度对TMB的影响



样本的突变丰度整体偏低时，更需要降低VAF避免纯度影响



C、E样本突变丰度偏低，在肿瘤纯度还没有特别低的时候，TMB值（黄线）开始降低，调整计算TMB的VAF（蓝色和绿色）能够更有效的还原TMB

结论：
 3、样本肿瘤突变丰度较低时（C、E样本），更需要降低计算TMB的VAF来还原肿瘤的TMB值
 — Unify VAF 5% TMB, 不考虑肿瘤纯度，统一采用5%
 ● 软件估算肿瘤纯度

TMB纯度偏离系数定义

$$\text{TMB纯度偏离系数} = \frac{\text{Purity_diff} * \text{TMB_diff}}{\text{F1_Score} * 100\% \text{ answer_TMB}} * \text{LOD} * \text{Depth}$$

影响TMB检出准确度的因素有：

- F1_score;
- 肿瘤细胞纯度带来的影响：Purity_diff=|40% TMB - 20% TMB|+|20% TMB-10% TMB|+|10% TMB - 5% TMB|
- Panel-TMB与WES-TMB的差异；TMB_diff = |40% TMB - 答案TMB|+|20% TMB - 答案TMB|+|10% TMB - 答案TMB|+|5% TMB-答案TMB|
- LOD：LOD = 声称LOD/实际LOD（实际LOD越低，挽回效果越好）
- 测序深度：Depth = 实际测序深度/声称测序深度（实际测序深度提升引起LOD降低）

部分单位由于上报结果不符合系数计算要求，不在后续结果中显示

Unpublished data 未经同意，不得使用传播

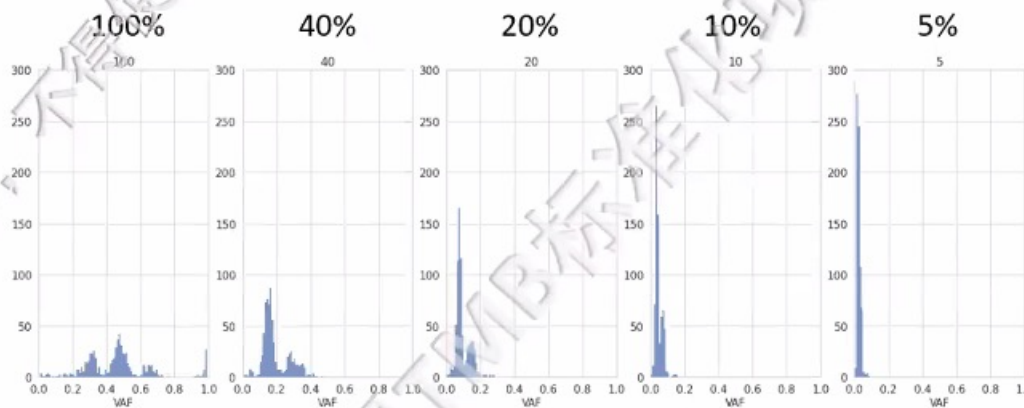
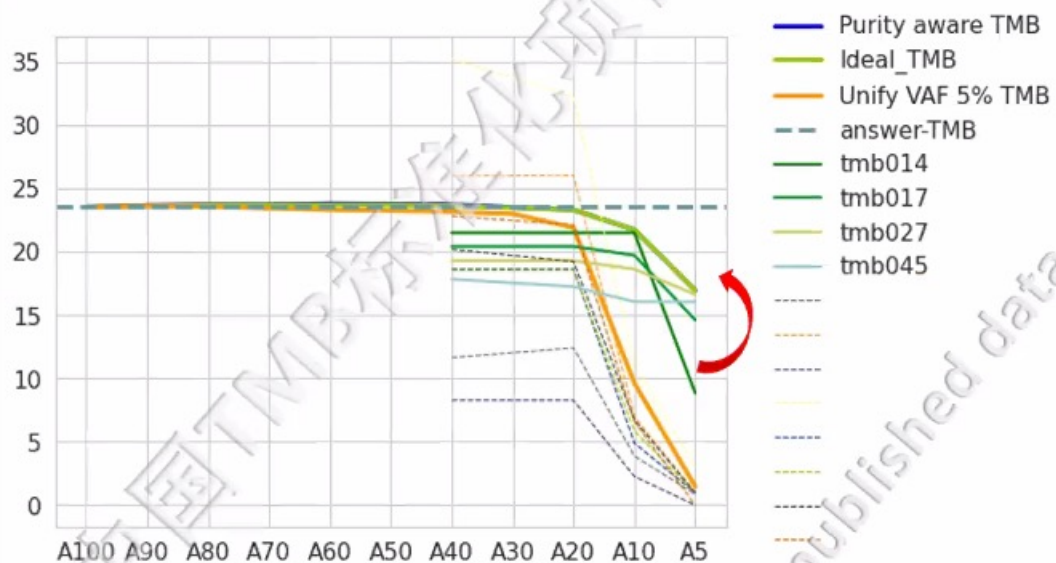
参与单位TMB结果受肿瘤纯度影响情况——样本A

ID	声称测序深度	实际测序深度	声称 SNV LOD	实际上报vcf的最低频率	计算TMB VAF	A40	A20	A10	A05	答案TMB	TMB纯度偏离系数
tmb027	500	1461	0.50%	1.00%	2.00%	19.22	19.28	18.61	16.62	24.047	4.190
tmb017	1600	1396	1.00%	0.70%	1.00%	20.4	20.4	19.7	14.6	24.047	6.876
tmb045	500	3035	0.40%	0.70%	1.00%	17.8	17.23	16.04	16.04	24.047	7.837
tmb014	1000	1838	2.00%	2.00%	2.00%	21.48	21.48	21.48	8.85	24.047	22.626
tmbxxx	500	1534	2.00%	1.80%	5.00%	8.27	8.27	2.26	0	24.047	93.436
tmbxxx	500	2256	2.00%	2.00%	5.00%	18.5	18.51	5.84	0.97	24.047	181.386
tmbxxx	500	2434	2.00%	2.00%	5.00%	18.6	18.6	4.89	0.98	24.047	199.443
tmbxxx	500	2512	2.00%	0.90%	5.00%	20.2	19.21	6.72	0.96	24.047	443.040
tmbxxx	500	1734	2.00%	0.70%	5.00%	26	26	7	1	24.047	459.684
tmbxxx	500	1020	1.00%	0.20%	5.00%	22.8	22.1	6.6	0	24.047	463.224
tmbxxx	250	1802	2.00%	1.30%	5.00%	35.2	32.18	11.06	3.02	24.047	825.211
tmbxxx	1000	4403	5.00%	0.40%	5.00%	11.6	12.4	3.88	0.86	24.047	1936.383

系数≤25

系数>25

参与单位TMB结果受肿瘤纯度影响情况——样本A



样本A中位点突变丰度呈正太分布

与虚线相比，实线更接近于answer TMB

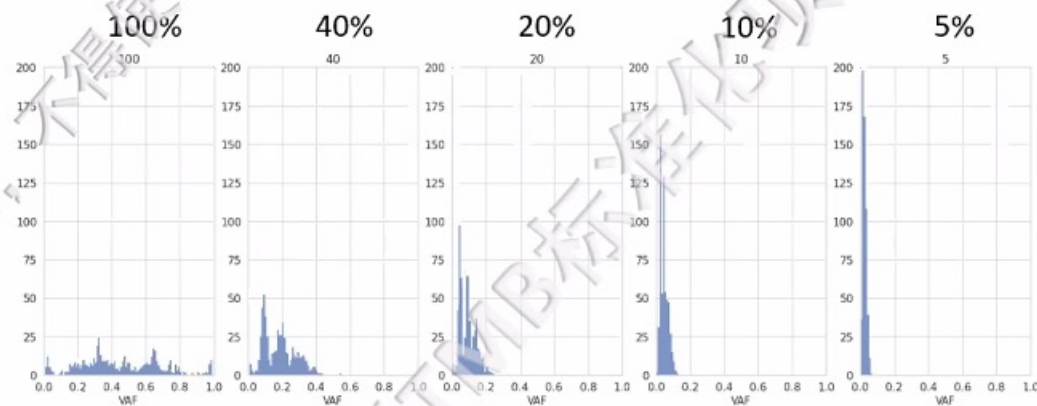
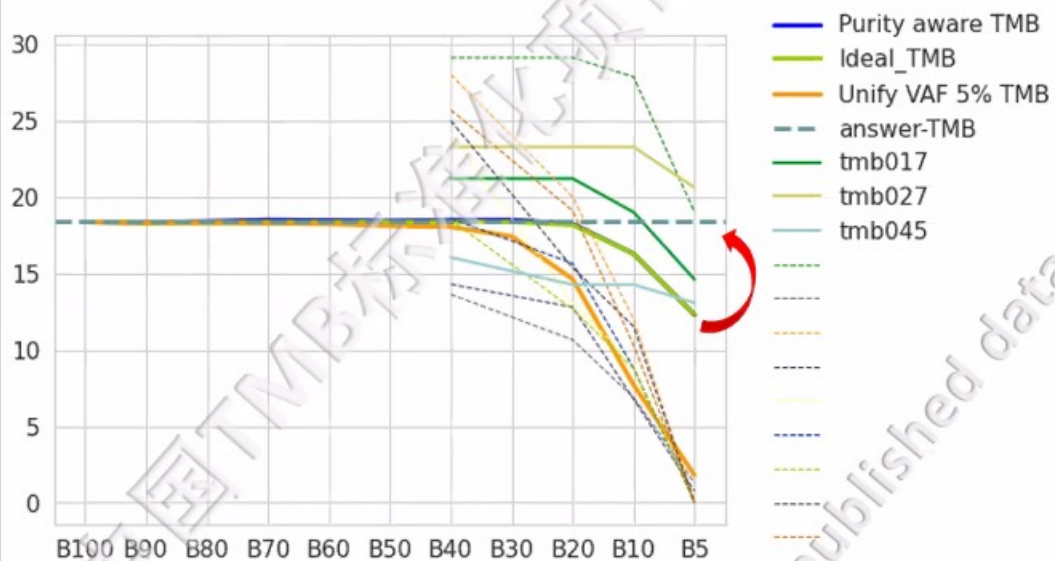
参与单位TMB结果受肿瘤纯度影响情况——样本B

ID	声称测序深度	实际测序深度	声称 SNV LOD	实际上报vcf的最低频率	计算TMB VAF	B40	B20	B10	B05	答案TMB	TMB纯度偏离系数
tmb027	500	1461	0.50%	1.00%	2.00%	23.27	23.27	23.27	20.61	19.167	3.227
tmb017	1600	1396	1.00%	0.70%	1.00%	21.20	21.20	19.00	14.60	19.167	4.096
tmb045	500	3035	0.40%	0.70%	1.00%	16.04	14.26	14.26	13.07	19.167	10.749
tmbxxx	1000	1838	2.00%	2.00%	2.00%	29.11	29.11	27.85	18.99	19.167	28.552
tmbxxx	500	1534	2.00%	1.80%	5.00%	14.29	12.79	6.77	0.75	19.167	104.331
tmbxxx	500	2434	2.00%	2.00%	5.00%	18.59	15.65	8.80	0.00	19.167	167.096
tmbxxx	500	2256	2.00%	2.00%	5.00%	18.50	12.66	8.76	0.00	19.167	168.461
tmbxxx	500	1020	1.00%	0.20%	5.00%	25.70	19.10	10.30	0.00	19.167	507.693
tmbxxx	500	1734	2.00%	0.70%	5.00%	28.00	20.00	12.00	0.00	19.167	528.487
tmbxxx	500	2512	2.00%	0.90%	5.00%	24.98	15.37	11.53	0.00	19.167	535.732
tmbxxx	250	1802	2.00%	1.30%	5.00%	24.13	13.07	9.05	0.00	19.167	587.986
tmbxxx	1000	4403	5.00%	0.40%	5.00%	13.63	10.65	6.87	1.28	19.167	1600.831

系数≤25

系数>25

参与单位TMB结果受肿瘤纯度影响情况——样本B



样本B中位点突变丰度呈正太分布

与虚线相比，实线更接近于answer TMB

Unpublished data 未经同意，不得使用传播

参与单位TMB结果受肿瘤纯度影响情况——样本D

ID	声称测序深度	实际测序深度	声称 SNV LOD	实际上报vcf的最低频率	计算TMB VAF	D40	D20	D10	D05	答案TMB	TMB纯度偏离系数
tmb027	500	1461	0.50%	1.00%	2.00%	9.97	9.97	9.97	9.31	7.416	1.438
tmb017	1600	1396	1.00%	0.70%	1.00%	7.30	6.60	5.80	3.60	7.416	4.293
tmb045	500	3035	0.40%	0.70%	1.00%	7.72	6.53	5.35	5.35	7.416	6.198
tmb014	1000	1838	2.00%	2.00%	2.00%	13.90	13.90	13.90	11.37	7.416	15.020
tmbxxx	500	2256	2.00%	2.00%	5.00%	11.68	11.68	5.84	1.95	7.416	97.064
tmbxxx	500	2434	2.00%	2.00%	5.00%	11.74	11.74	6.85	1.96	7.416	99.145
tmbxxx	250	1802	2.00%	1.30%	5.00%	9.05	9.05	6.03	1.01	7.416	138.864
tmbxxx	500	1534	2.00%	1.80%	5.00%	15.04	12.79	11.28	1.50	7.416	145.986
tmbxxx	500	1734	2.00%	0.70%	5.00%	14.00	13.00	8.00	2.00	7.416	295.394
tmbxxx	500	1020	1.00%	0.20%	5.00%	20.60	19.10	13.20	2.90	7.416	917.550
tmbxxx	1000	4403	5.00%	0.40%	5.00%	11.08	10.25	8.11	0.87	7.416	1062.071
tmbxxx	500	2512	2.00%	0.90%	5.00%	24.98	24.02	17.29	2.88	7.416	1634.053

系数≤25

系数>25

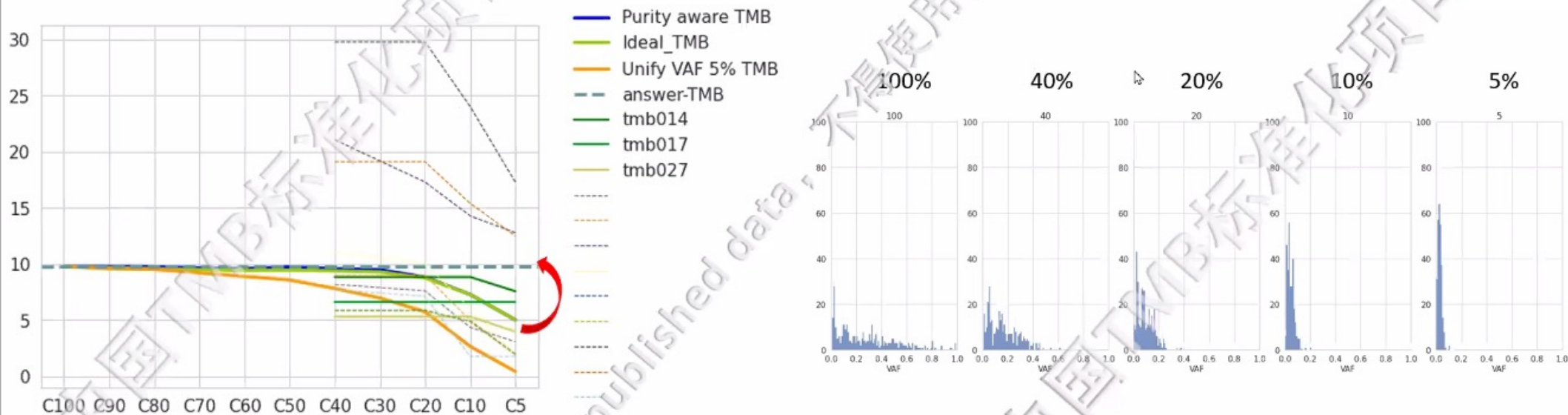
参与单位TMB结果受肿瘤纯度影响情况——样本C

ID	声称测序深度	实际测序深度	声称 SNV LOD	实际上报vcf的最低频率	计算TMB VAF	C40	C20	C10	C05	答案TMB	TMB纯度偏离系数
tmb017	1600	1396	1.00%	0.70%	1.00%	6.60	6.60	6.60	6.60	10.37	0.000
tmb014	1000	1838	2.00%	2.00%	2.00%	8.85	8.85	8.85	7.58	10.37	1.693
tmb027	500	1461	0.50%	1.00%	2.00%	5.32	5.32	5.32	3.99	10.37	4.669
tmbxxx	500	2256	2.00%	2.00%	5.00%	5.85	5.85	4.87	1.95	10.37	40.991
tmbxxx	500	2434	2.00%	2.00%	5.00%	5.87	5.88	4.90	1.96	10.37	44.274
tmbxxx	500	3035	0.40%	0.70%	1.00%	7.72	7.13	1.78	1.78	10.37	48.147
tmbxxx	500	1534	2.00%	1.80%	5.00%	21.06	17.30	14.29	12.79	10.37	67.082
tmbxxx	500	1734	2.00%	0.70%	5.00%	9.00	9.00	5.00	2.00	10.37	111.792
tmbxxx	250	1802	2.00%	1.30%	5.00%	11.06	10.06	6.03	2.01	10.37	138.405
tmbxxx	500	1020	1.00%	0.20%	5.00%	19.10	19.10	15.40	12.50	10.37	171.306
tmbxxx	1000	4403	5.00%	0.40%	5.00%	8.19	7.62	4.37	3.09	10.37	502.961
tmbxxx	500	2512	2.00%	0.90%	5.00%	29.78	29.78	24.02	17.29	10.37	807.491

系数≤25

系数>25

参与单位TMB结果受肿瘤纯度影响情况——样本C



样本C中位点突变丰度呈左侧分布

与虚线相比，实线更接近于answer TMB

参与单位TMB结果受肿瘤纯度影响情况——样本E

ID	声称测序深度	实际测序深度	声称 SNV LOD	实际上报vcf的最低频率	计算TMB VAF	D40	D20	D10	D05	答案TMB	TMB纯度偏离系数
tmb027	500	1461	0.50%	1.00%	2.00%	5.32	5.32	5.32	2.66	4.75	3.593
tmb014	1000	1838	2.00%	2.00%	2.00%	6.32	5.05	3.79	3.79	4.75	3.793
tmb017	1600	1396	1.00%	0.70%	1.00%	8.50	7.30	6.60	3.60	4.75	14.207
tmb045	500	3035	0.40%	0.70%	1.00%	7.72	6.53	5.94	4.16	4.75	17.831
tmbxxx	500	1534	2.00%	1.80%	5.00%	3.76	3.01	1.50	0.00	4.75	29.819
tmbxxx	500	2256	2.00%	2.00%	5.00%	6.82	6.82	3.90	0.97	4.75	51.230
tmbxxx	500	2434	2.00%	2.00%	5.00%	6.85	6.85	2.94	0.98	4.75	62.020
tmbxxx	250	1802	2.00%	1.30%	5.00%	5.03	4.02	3.02	1.01	4.75	63.547
tmbxxx	500	1734	2.00%	0.70%	5.00%	8.00	6.00	3.00	3.00	4.75	84.625
tmbxxx	500	2512	2.00%	0.90%	5.00%	6.72	6.72	4.80	0.96	4.75	106.500
tmbxxx	500	1020	1.00%	0.20%	5.00%	8.10	7.40	3.70	0.70	4.75	189.051
tmbxxx	1000	4403	5.00%	0.40%	5.00%	3.88	4.30	1.72	1.71	4.75	262.996

系数≤25

系数>25

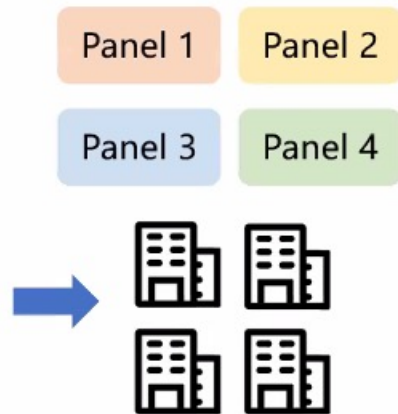
项目第一阶段时间线



C、临床样本评估



- 50例临床样本;
- NCC统一提取DNA和质控
- 200ng的肿瘤样本和癌旁对照
- 部分样本提供HE切片和白片
- 高深度WES做TMB参考标准



DNA下发各参与单位, 各参与单位使用公司产品测序

建库测序



上报测序数据fastq/bam、vcf、解读报告 (14-20天上传vcf和解读报告, 30天内上传fastq/bam)



C、临床样本进行多维度挑选

临床样本的选择从多个维度出发，全面评估参与单位产品性能

评估突变检出 (SNV、indel、CNV、Fusion、MSI)

A

B

评估TMB定性定量能力

评估解读能力

C

D

肿瘤纯度、样本质量

覆盖常见癌种

E

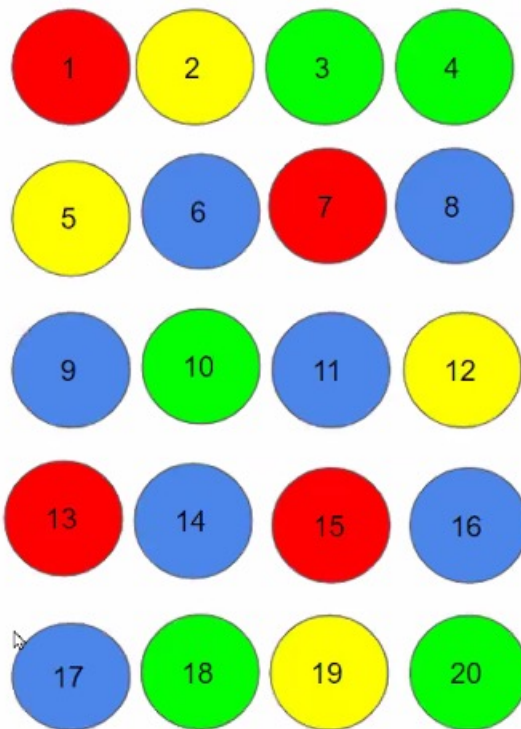
F

紧贴临床治疗方案

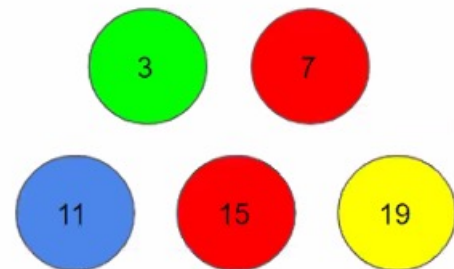
C、目前已完成回顾性临床样本筛选

- 筛选出超过100例临床样本，满足下列评估：

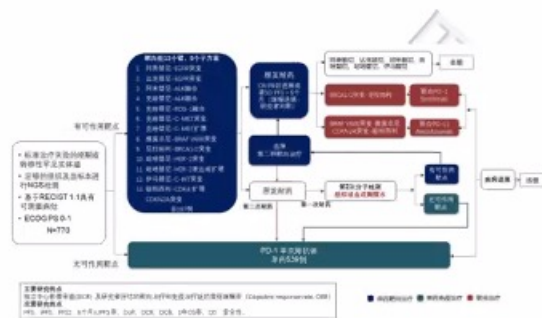
- ◆ 评估Panel在**不同肿瘤细胞占比**下突变检出能力
- ◆ 评估Panel在**不同的DNA样本质量**下突变检出能力
- ◆ 评估不同**突变丰度位点检出准确度**
- ◆ Panel检测与WES检测**一致度**
 - TMB定量一致性
- ◆ Panel检测**MSI**与CE检测一致性分析
- ◆ Tier分级**解读**准确性
- ◆ 临检**速度**（参考）
- ◆ Panel预测新生抗原能力（调研）



样本按照规则
打乱顺序发放



第二阶段——TMB临床获益基线的划定



Platform-罕见癌种



100*10个常见癌种
 回顾性/前瞻性
 真实世界临床样本

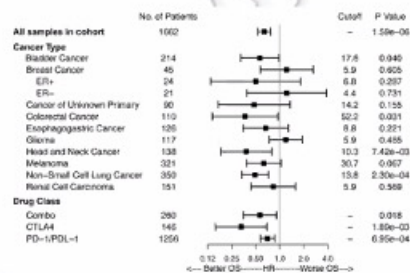
大Panel测序



WES测序



上传Cancer Gene Testing Analysis平台



不同癌种TMB中国人群基线



黄杰

CONTENTS

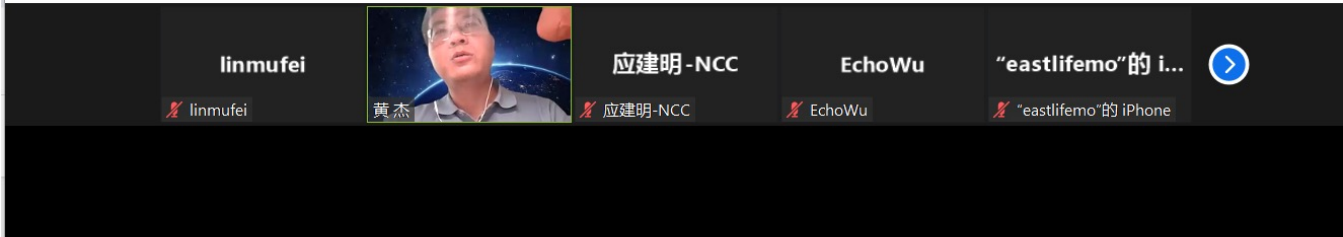
目录

1

项目1B阶段方案与参考品答案集

2

项目1A-1B阶段标准化结果分享



医用高通量测序正式进入标准化时代



5月28日，国家药监局批准医疗器械标准管理中心，正式筹建**医用高通量测序标准化技术归口单位**。此举意味着**医用高通量测序行业**将会迎来标准化时代！



索引号	FGWJ-2021-10001	主题分类	法规文件 / 规范性文件
标题	国家药监局综合司关于同意筹建中医器械标准化技术归口单位和 医用高通量测序标准化技术归口单位的复函		
发布日期	2021-05-28		

国家药监局综合司关于同意筹建中医器械标准化技术归口单位和 医用高通量测序标准化技术归口单位的复函

药监综械注函〔2021〕311号



发布时间: 2021-05-28

中国TMB标准化项目1B阶段为细胞系参考品评估

项目立项与报名		第一阶段		第二阶段
1. 伦理申请	A: 模拟数据评估	B: 细胞系参考品评估	C: 临床样本评估	D: 分癌种启动 TMB cut-off值构建 临床研究
2. 人遗备案				
3. 注册临床试验	应用大样本量 WES数据评估参与单位TMB产品与WES拟合情况, 统一方法建立各产品区域校正参数	通过突变结果统计分析及湿实验验证建立标准品答案集, 评估各产品位点检测准确度、TMB一致性等指标	评估各单位位点检出准确度、TMB一致性和解读准确度, 验证统一构建的校正参数, 建立TMB标准化评估体系	选择8-10个癌种, 收集带有ICI用药随访数据样本, 进行TMB检测及标准化的评估, 建立药物疗效cut-off值
4. 与各参与方签订项目合作协议				
5. 各项准备工作				

B、参考品评估

中检院提供参考品 DNA
(500ng的肿瘤参考品和对照
参考品)，同时使用高深度
WES做TMB参考标准



Cell line参考品

WES

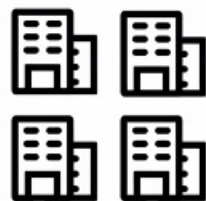
共计25个参考品

Panel 1

Panel 2

Panel 3

Panel 4



DNA下发各参与单
位，各参与单位使
用公司产品测序

建库测序



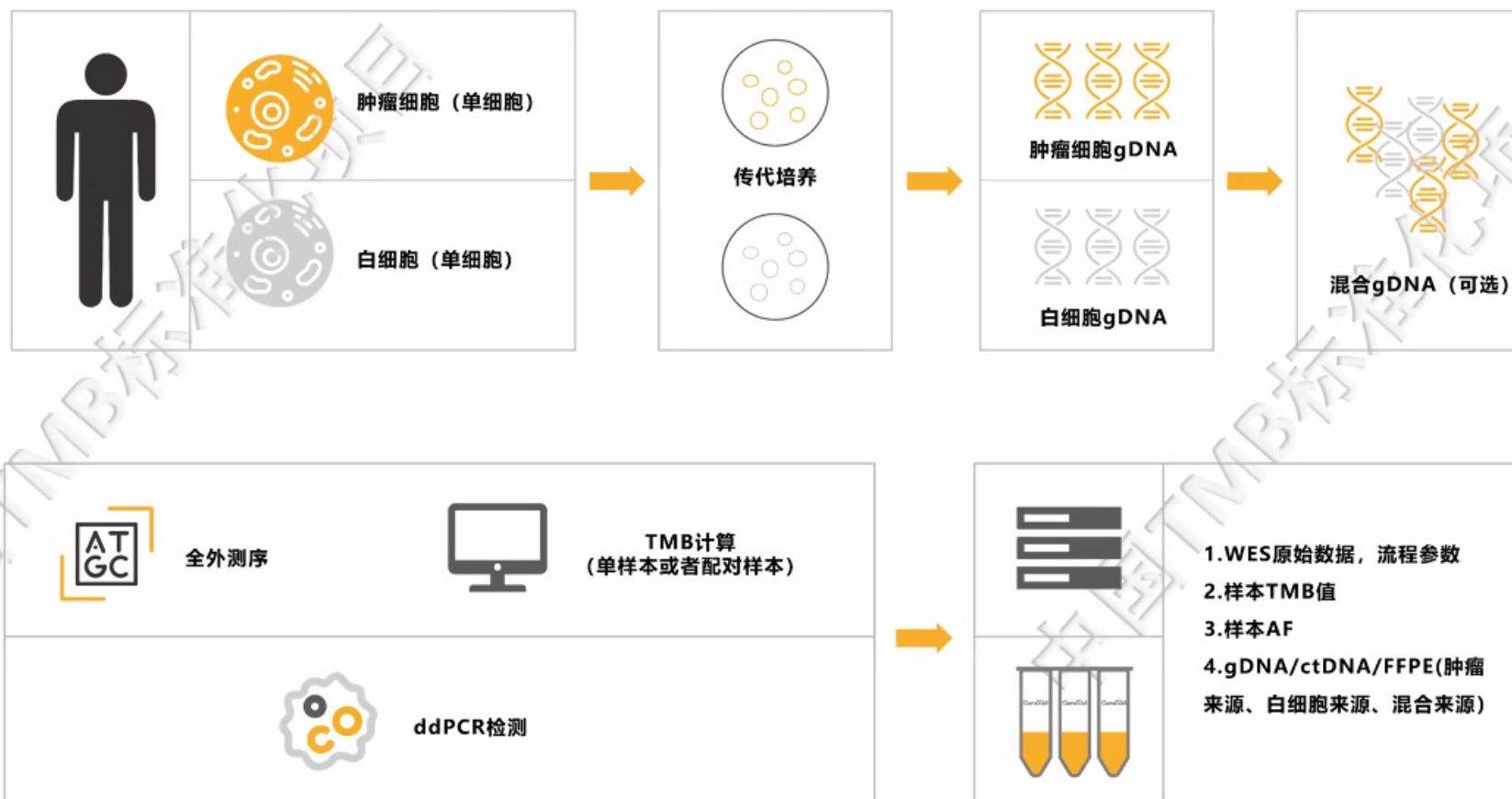
Cancer Gene Testing Analysis (CGTA)



上报测序数据fastq/bam、vcf (14
天上传vcf, 30内上传fastq/bam)



TMB参考品制备流程



确定最低肿瘤细胞含量：与肿瘤突变丰度高度相关

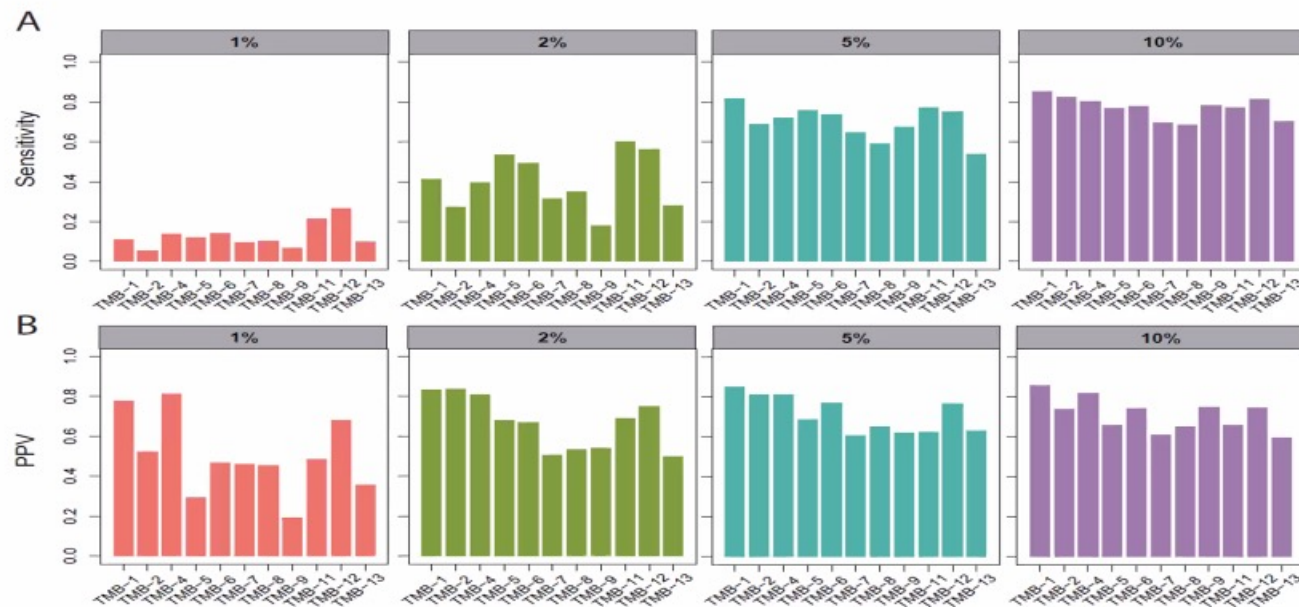


Figure S2. WES demonstrates limited variant detection proficiency at low VAF. A dilution series of each standard sample was produced at 1%, 2%, 5%, and 10% mass-by-mass ratio with the respective matched normal control. Each step of the gradient was sequenced in duplicates. The union of variant calls among the two replicates was used for sensitivity and PPV analysis. A) Variant detection sensitivity at each dilution compared to the variant calls in the undiluted sample. B) Variant detection PPV at each dilution compared to the variant calls in the undiluted sample.

确定稳定检出的突变比例

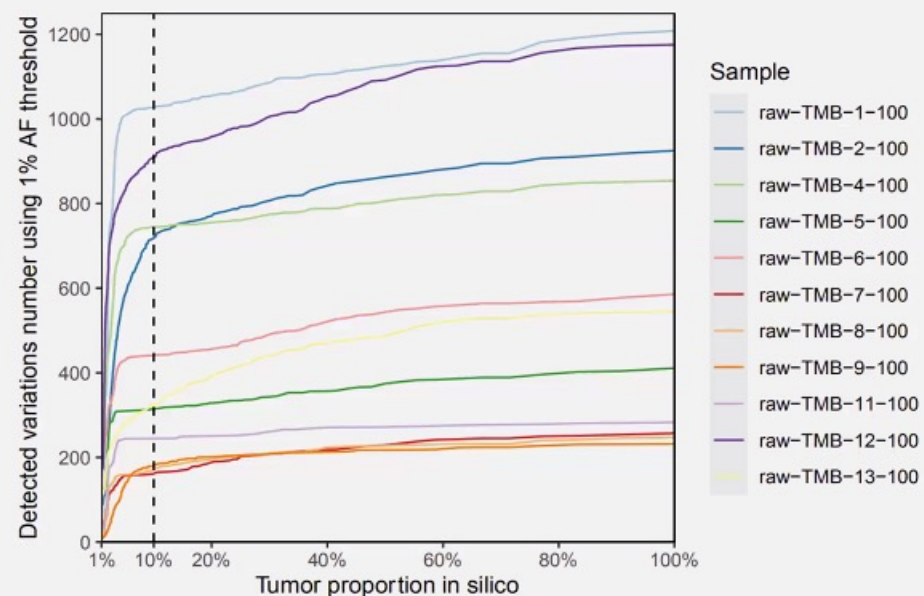
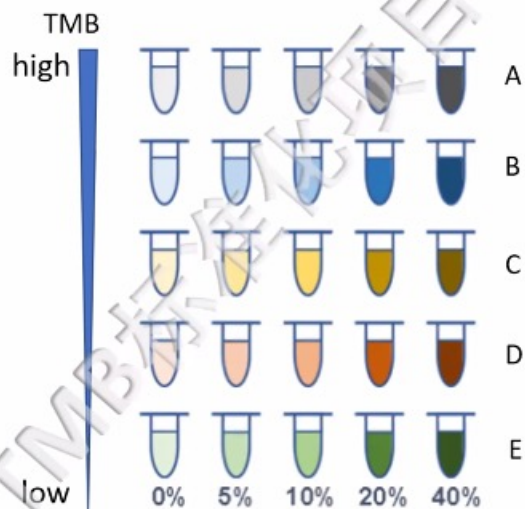


Figure 2. Detected variations number using 1% AF threshold increases with tumor proportion from 1% to 100% *in silico*.

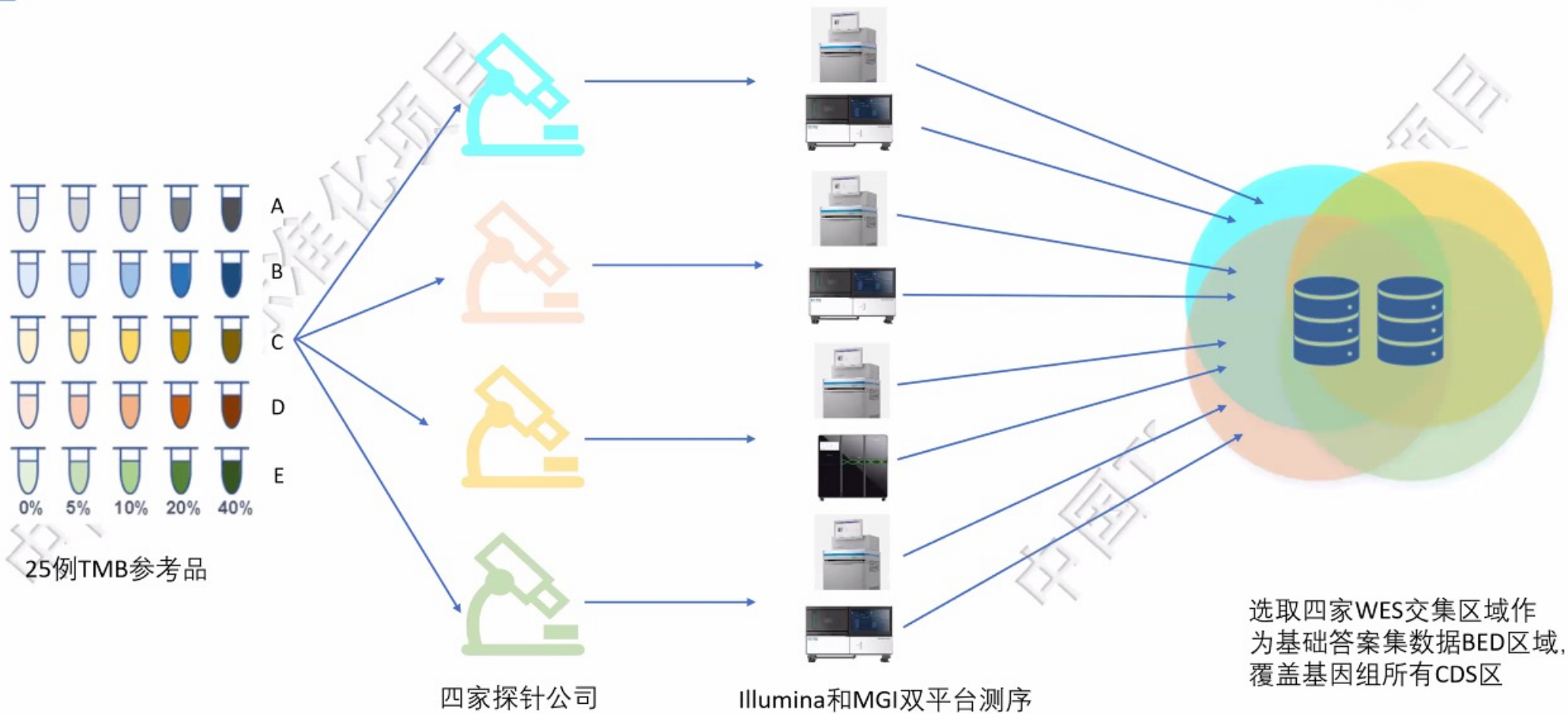
项目使用不同肿瘤细胞比例、不同TMB值的参考品



每套TMB标准品包含5组共25支，每组5支。TMB标准品均由肿瘤细胞和配对的正常细胞混合而成，其中肿瘤细胞比例分别为40%、20%、10%、5%和0%。

No.	purity	names	100%浓度 标品TMB
1	40%	A40	24.04738
2	20%	A20	24.04738
3	10%	A10	24.04738
4	5%	A05	24.04738
5	0%	A0	0
6	40%	B40	19.16727
7	20%	B20	19.16727
8	10%	B10	19.16727
9	5%	B05	19.16727
10	0%	B0	0
11	40%	C40	10.37023
12	20%	C20	10.37023
13	10%	C10	10.37023
14	5%	C05	10.37023
15	0%	C0	0
16	40%	D40	7.416482
17	20%	D20	7.416482
18	10%	D10	7.416482
19	5%	D05	7.416482
20	0%	D0	0
21	40%	E40	4.751685
22	20%	E20	4.751685
23	10%	E10	4.751685
24	5%	E05	4.751685
25	0%	E0	0

标品采用4家WES探针，双平台测序共生成8套数据集



WES探针双平台WES测序质量良好，满足答案集构建的要求



Company	A公司		B公司		C公司		D公司	
platform	ILMN (PE150)	MGI (PE100)	ILMN (PE150)	MGI (PE150)	ILMN (PE150)	MGI (PE150)	ILMN (PE150)	MGI (PE150)
Q30	94.02±0.16	93.34±0.14	94.91±0.24	85.74±0.77	94.76±0.30	90.67±0.52	93.60±0.55	87.48±0.64
CQ30	0.85±0.01	0.80±0.01	0.84±0.01	0.86±0.01	0.84±0.01	0.86±0.01	0.83±0.01	0.85±0.01
Duplicate (%)	25.09±3.62	11.77±0.84	25.47±0.95	6.41±0.91	26.96±5.06	9.13±1.23	28.03±4.34	9.11±1.57
Capture_efficiency_dup(%)	50.03	52.41	52.24	51.68	53.70	54.25	54.17	48.08
0.5X_Mean_Covered(%)	94.19	94.51	93.12	94.02	94.15	93.64	90.01	92.86
AverageSeqDepth_Raw	1268.4	1139.8	1226.3	886.4	794.7	750.9	1079.3	956.5
AverageSeqDepth_Dup	932.8	1002.9	905.7	823.6	568.2	676.9	792.8	890.1
Fold80	1.46	1.45	1.48	1.37	1.40	1.39	1.59	1.47

1、四家公司测序质量良好，均满足构建答案集的要求；

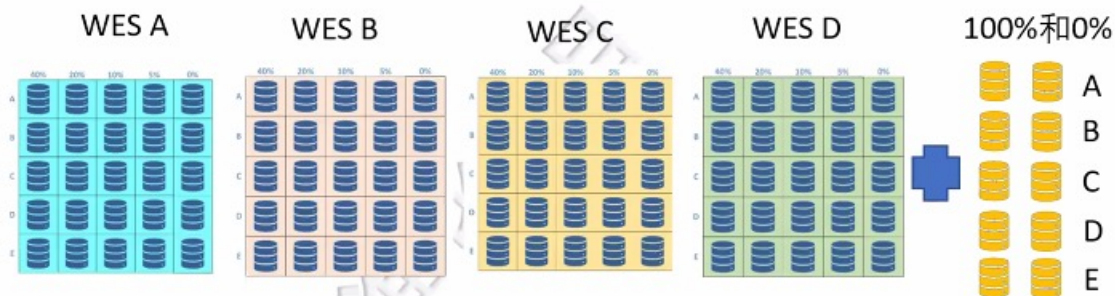
2、MGI的duplicate优于ILMN，ILMN的Q30略优于MGI

	ILMN	MGI
Q30	94.3±0.64	89.3±2.98
CQ30	0.842±0.028	0.838±0.01
Duplicate (%)	26.4±3.93	9.07±2.22

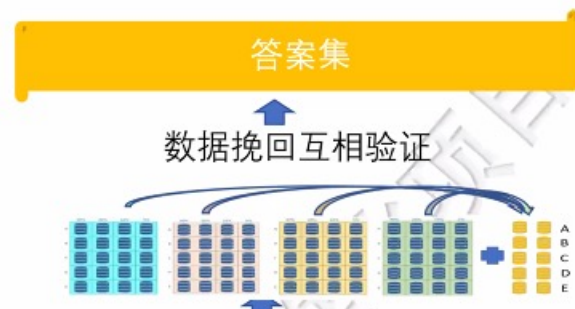
关注测到的每个碱基的准确度，Q30代表测序正确率≥99.9%的碱基占比。

Consensus Quality 30 (CQ30)
定义：关注每个有突变信号的共有序列碱基的准确度，质量值CQ30代表测序正确率≥99.9%的共有序列碱基占比。

本次试验答案集构建流程



ILMN + MGI 双平台测序数据



人工校验 + 实验校验

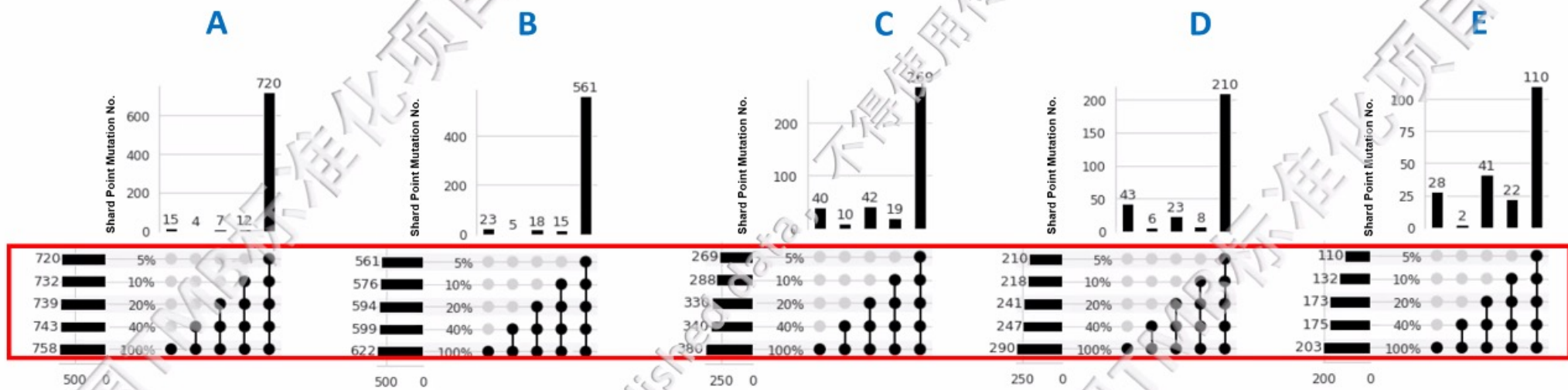
- 低浓度样本在高浓度没有检出，需要人工IGV位点确认
- 可疑的位点，经菁良实验验证

突变需同时满足如下条件

- SNV: 6个软件有3个支持;
- Indel: 4个软件中至少有2个支持;
- 100%纯度中有支持;
- 8套数据中至少2套支持;
- Illumina和MGI平台都有支持;
- Dbsnp及ExAC MAF≥0.01人群过滤

答案集经过**专家组共同设计和认证**，根据答案构建流程筛选得到最终每个样本答案的突变位点。

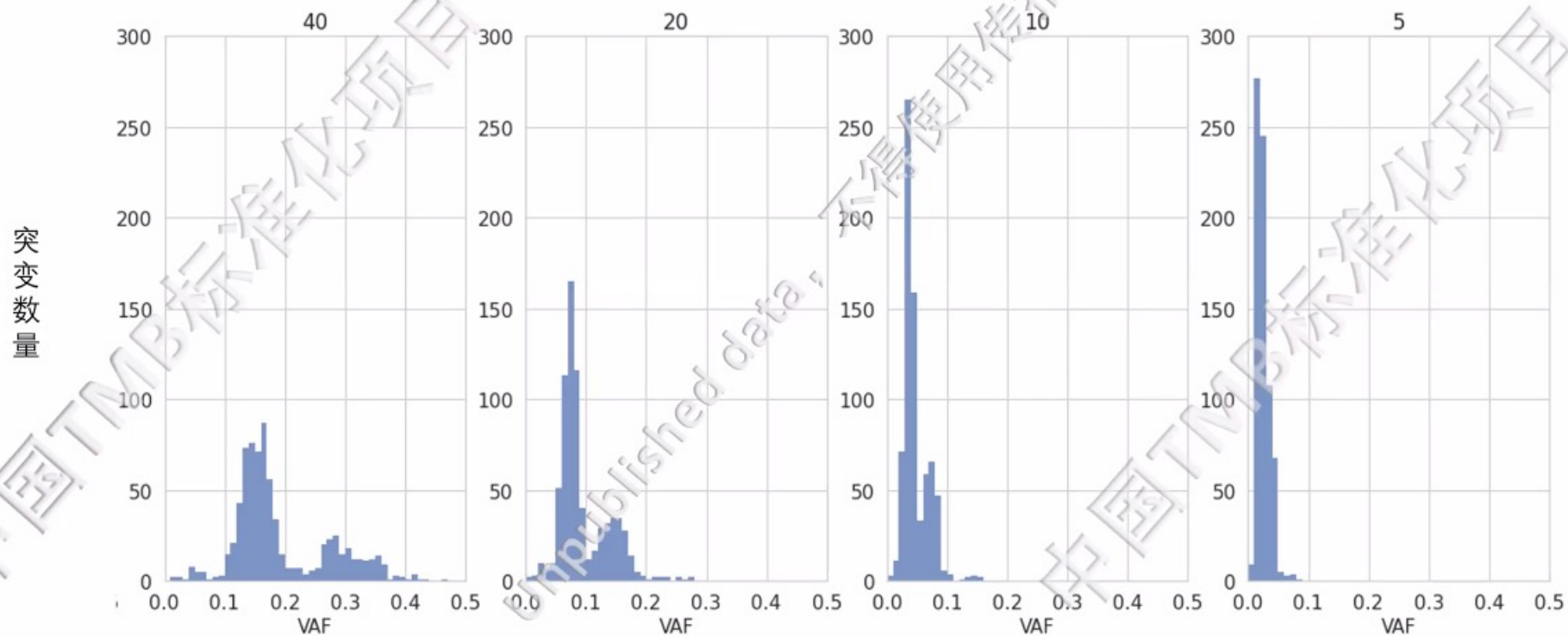
不同比例的参考品突变统计



随着肿瘤细胞比例的下降，可检出肿瘤突变数量降低

(答案集过滤人群SNPs、同义突变，突变位于CDs区域，突变丰度 $\geq 1\%$)

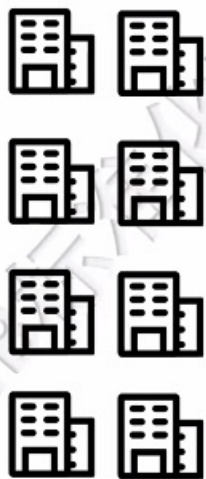
参考品中SNV和Indel的VAF分布(参考品A)



2

项目1A-1B阶段标准化结果分享

A、模拟数据预评估——提交产品参数



参与单位

提交产品信息

- 专注的癌种;
- Bed文件;
- LOD;
- 有效测序深度;
- 检测突变类型;
- 测序数据量;
- 报告时间;
- 其他指标;

CANCER 抗癌登月
www.0cancer.cn



Cancer Gene Testing Analysis (CGTA)

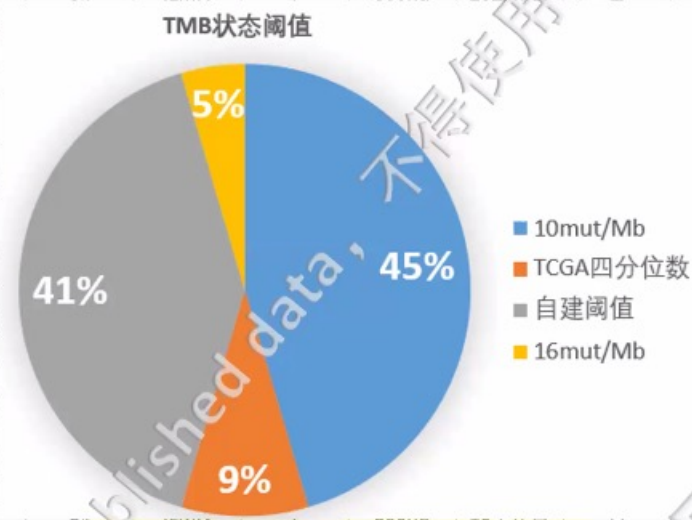
参与单位上报的性能统计



序号	TMB状态 阈值	SNV检出 LOD	InDel检出 LOD	CNV Gain检出 LOD	CNV Loss检出 LOD	重排/融 合检出 LOD	TMB肿瘤 细胞占比 要求	测序平台 缩写	MSI检出 肿瘤细胞 占比要求	对肿瘤样 本DNA上 样量要求	报告时间	是否需要 RNA检测 Fusion	对肿瘤样 本RNA上 样量要求	去重前测 序深度	去重后测 序深度	肿瘤样本 下机测序 数据量	对照样本 下机测序 数据量	是否UMI
1	10	1%	1%	.	.	.	5%	ILMN	.	500ng	10工作日	否	.	2000	1000	6Gb	1Gb	UMI
2	10	2%	2%	3.5copies	1copies	8%	10%	ILMN	10%	200ng	8个自然 日	否	.	3500X	1000X	9G	3G	.
3	自建	5%	5%	.	.	.	20%	ILMN	.	300ng	10工作日	否	.	2000	1000	8Gb	0.8Gb	.
4	TCGA	1%	1%	.	.	.	5%	ILMN	.	100ng	10工作日	否	.	3000	1600	8Gb	1.5Gb	.
5	10	5%	5%	.	.	.	20%	ILMN	.	50ng	10-15工 作日	否	.	750	500	6Gb	1.5Gb	.
6	自建	2%	2%	.	.	.	20%	MGI	.	100ng	7工作日	否	.	1500	500	5Gb	1.2Gb	.
7	10	2%	5%	3copies	.	5%	20%	ILMN	20%	200ng	10工作日	否	.	1000	500	8Gb	2Gb	.
8	10	2%	2%	.	.	.	20%	ILMN	.	250ng	8-10工作 日	否	.	2000	500	7Gb	1Gb	.
9	自建	2%	2%	.	.	.	20%	ILMN	.	200ng	7工作日	否	.	1000	250	7.5Gb	7.5Gb	.
10	自建	1%	1%	.	.	.	50%	MGI	.	100ng	7工作日	否	.	3000	2500	12Gb	3Gb	.
11	自建	0.50%	0.50%	4copies	.	1%	20%	ILMN	20%	250ng	10工作日	否	.	2000	500	8Gb	3Gb	.
12	10	5%	8%	6copies	.	.	20%	ILMN	20%	50ng	10工作日	否	.	500	500	4Gb	0Gb	.
13	10	1%	1%	.	.	.	20%	ILMN	.	2ng	15工作日	否	.	2000	450	4Gb	2Gb	.
14	10	5%	5%	4copies	.	5%	20%	ILMN	20%	200ng	12工作日	否	.	1500	700	15Gb	3Gb	.
15	无	5%	10%	.	.	.	20%	ION	.	20ng	5工作日	否	.	1500	1500	4Gb	0Gb	.
16	自建	2%	3%	3copies	1.5copies	1%	20%	MGI	20%	100- 200ng	8工作日	否	.	1000	500	4.5Gb	1.5Gb	.
17	10	1%	1%	.	.	.	5%	ILMN	.	500ng	10工作日	否	.	2000	1000	6Gb	1Gb	UMI
18	自建	2%	3%	3copies	1.5copies	1%	20%	ILMN	20%	20-560ng	8工作日	是	50-300ng	2000	500	5Gb	1.32Gb	.
19	自建	2%	5%	3copies	.	5%	20%	ILMN	20%	200ng	10工作日	否	.	1000	500	8Gb	2Gb	.
20	自建	3%	3%	.	.	.	20%	MGI	.	200ng	7工作日	否	.	1200	1200	22Gb	3Gb	.
21	16	1%	1%	.	.	.	20%	ILMN	.	200ng	8工作日	否	.	600	500	6Gb	2Gb	.
22	TCGA	0.40%	0.40%	.	.	.	10%	ILMN	.	500ng	7工作日	否	.	1000	500	10Gb	2Gb	.
23	10	5%	5%	4copies	.	5%	20%	ILMN	20%	200ng	12工作日	否	.	1500	700	15Gb	3Gb	.

参与单位上报参数差异明显需要进行标准化

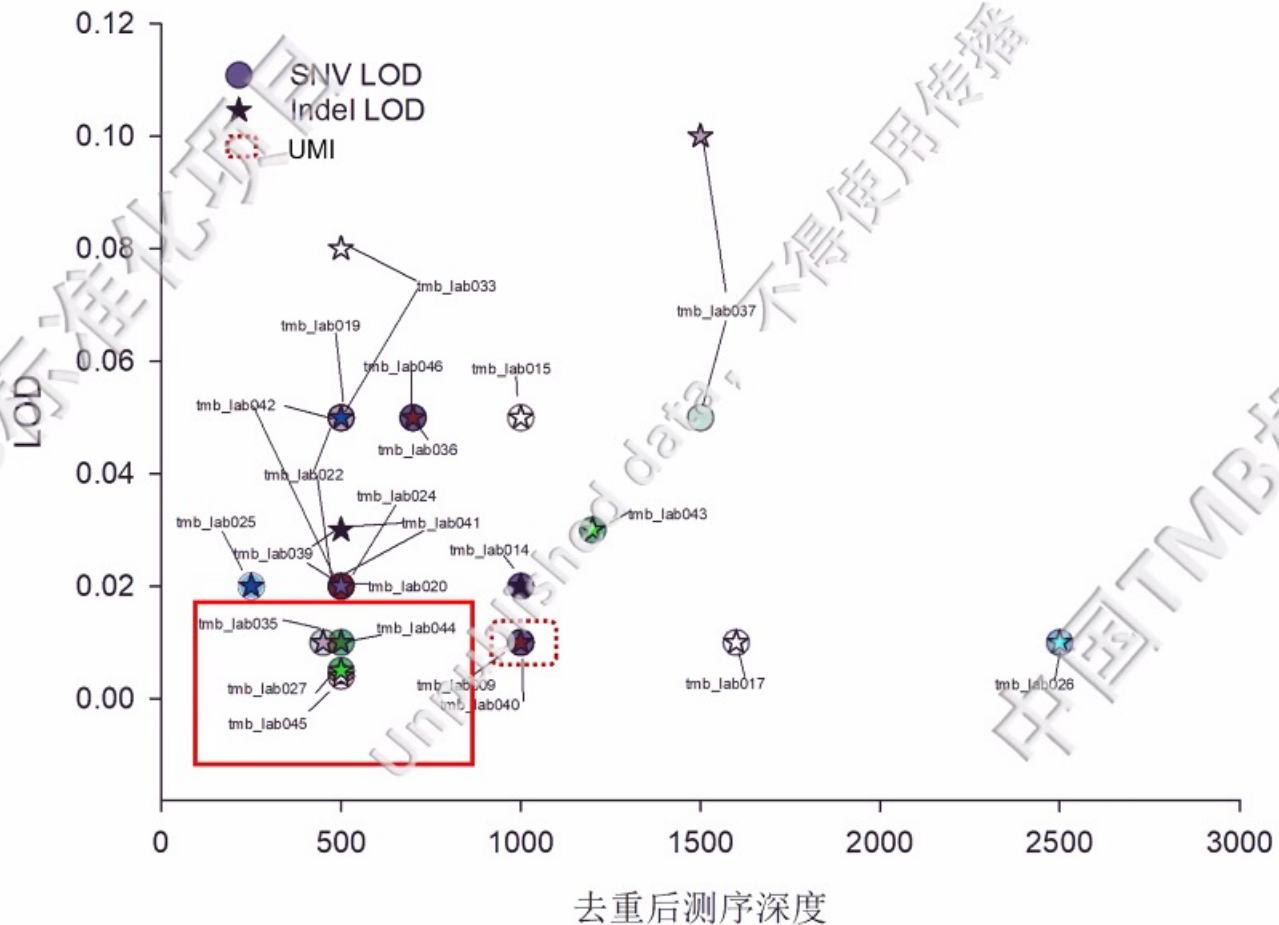
序号	TMB状态阈值	SNV检出LOD	InDel检出LOD	CNV Gain检出LOD	CNV Loss检出LOD	重排/融合检出LOD	TMB肿瘤细胞占比要求	测序平台缩写	MSI检出肿瘤细胞占比要求	对肿瘤样本DNA上样量要求	报告时间	是否需要RNA检测Eusion	对肿瘤样本RNA上样量要求	去重前测序深度	去重后测序深度	肿瘤样本下机测序数据量	对照样本下机测序数据量	是否UMI
1	10	1%	1%	.	.	.	5%	ILMN	.	500na	10工作日	否	.	2000	1000	6Gb	1Gb	UMI
2	10	2%	2%	3.5copies	1copies	8%								3500X	1000X	9G	3G	.
3	自建	5%	5%	.	.	.								2000	1000	8Gb	0.8Gb	.
4	TCGA	1%	1%	.	.	.								3000	1600	8Gb	1.5Gb	.
5	10	5%	5%	.	.	.								750	500	6Gb	1.5Gb	.
6	自建	2%	2%	.	.	.								1500	500	5Gb	1.2Gb	.
7	10	2%	5%	3copies	.	5%								1000	500	8Gb	2Gb	.
8	10	2%	2%	.	.	.								2000	500	7Gb	1Gb	.
9	自建	2%	2%	.	.	.								1000	250	7.5Gb	7.5Gb	.
10	自建	1%	1%	.	.	.								3000	2500	12Gb	3Gb	.
11	自建	0.50%	0.50%	4copies	.	1%								2000	500	8Gb	3Gb	.
12	10	5%	8%	6copies	.	.								102500	250	4Gb	0Gb	.
13	10	1%	1%	.	.	.								2000	450	4Gb	2Gb	.
14	10	5%	5%	4copies	.	5%								1500	700	15Gb	3Gb	.
15	无	5%	10%	.	.	.								1500	1500	4Gb	0Gb	.
16	自建	2%	3%	3copies	1.5copies	1%								1000	500	4.5Gb	1.5Gb	.
17	10	1%	1%	.	.	.								2000	1000	6Gb	1Gb	UMI
18	自建	2%	3%	3copies	1.5copies	1%	20%					是	50-300ng	2000	500	5Gb	1.32Gb	.
19	自建	2%	5%	3copies	.	5.00%	20%	ILMN	20%	200ng	10工作日	否	.	1000	500	8Gb	2Gb	.
20	自建	3%	3%	.	.	.	20%	MGI	.	200ng	7工作日	否	.	1200	1200	22Gb	3Gb	.
21	16	1%	1%	.	.	.	20%	ILMN	.	200ng	8工作日	否	.	600	500	6Gb	2Gb	.
22	TCGA	0.40%	0.40%	.	.	.	10%	ILMN	.	500ng	7工作日	否	.	1000	500	10Gb	2Gb	.
23	10	5%	5%	4copies	.	5%	20%	ILMN	20%	200ng	12工作日	否	.	1500	700	15Gb	3Gb	.



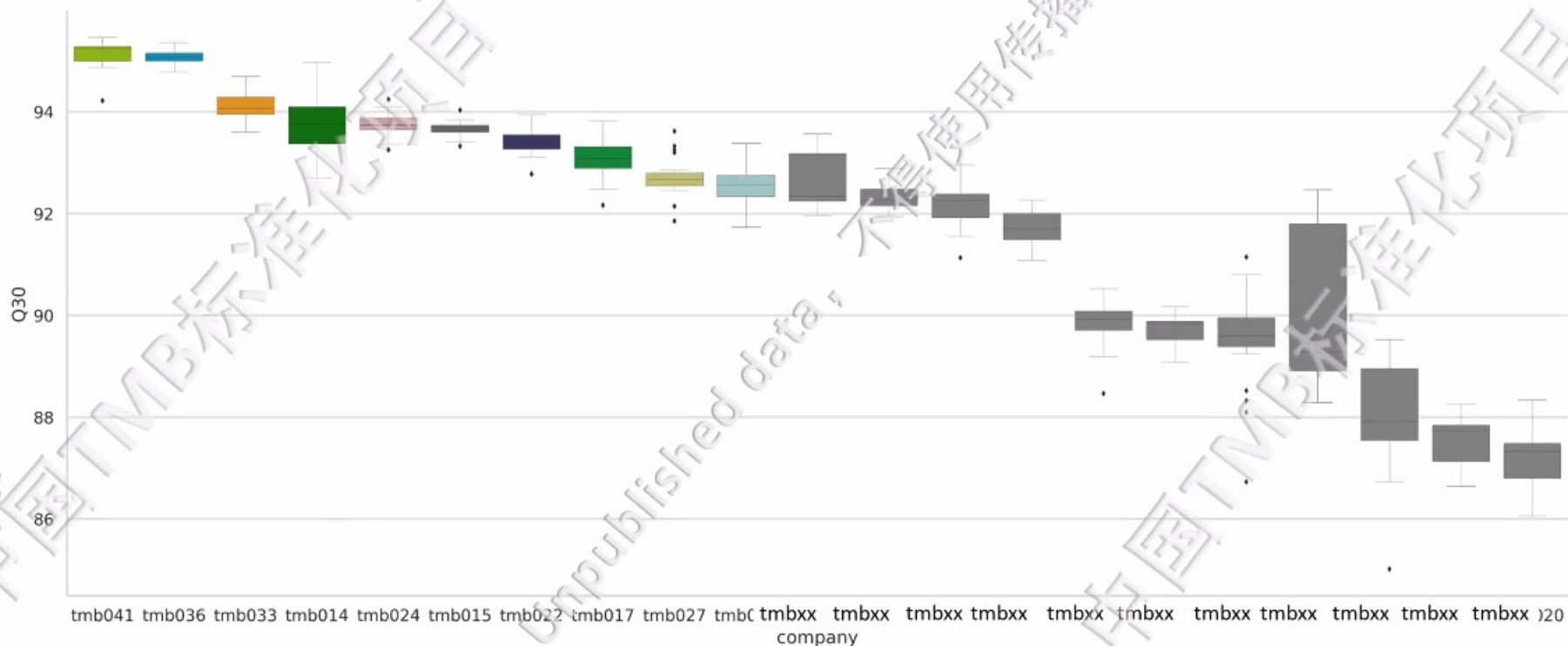
参与单位划定TMB-H阈值的标准

各参与单位测序深度、位点检出LOD、TMB状态阈值划定等均有所不同，亟需对TMB检测制定标准

突变检出LOD与去重后测序深度分布散点图

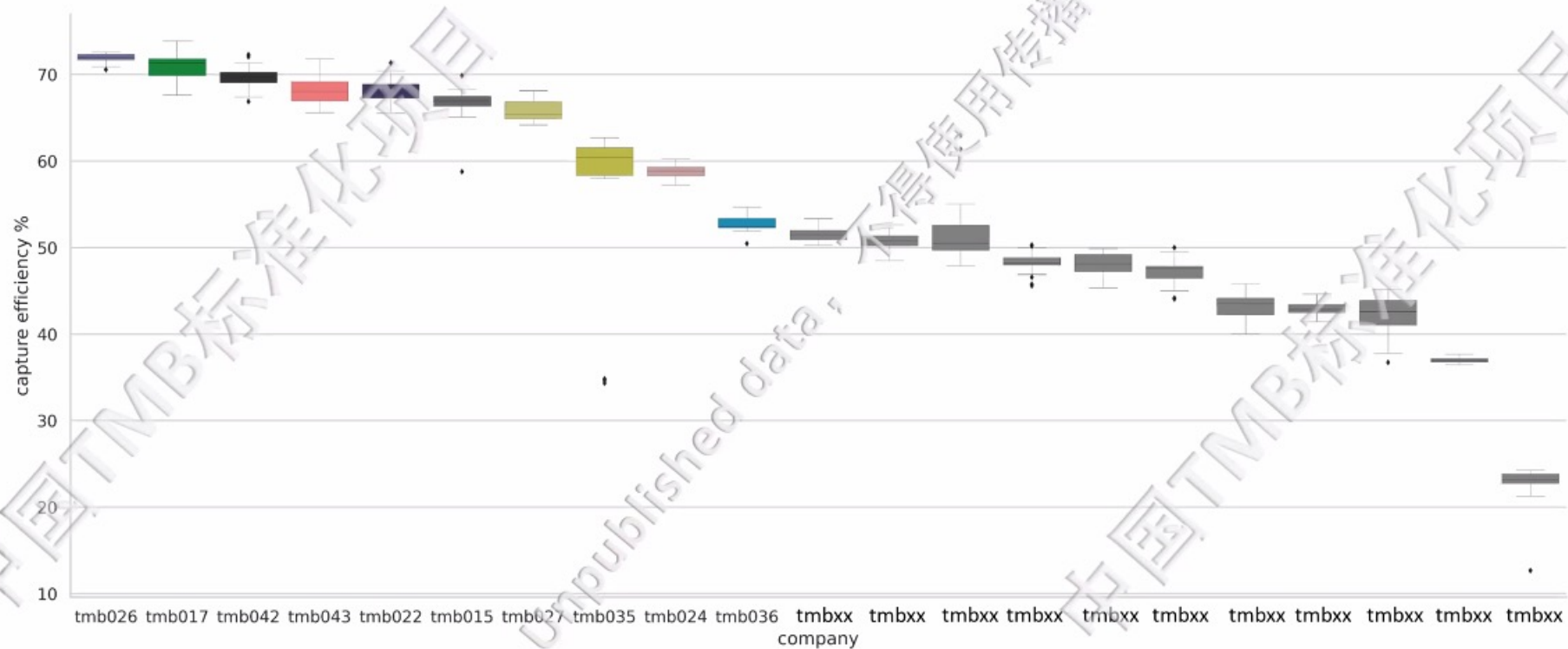


参与单位质控数据——Q30



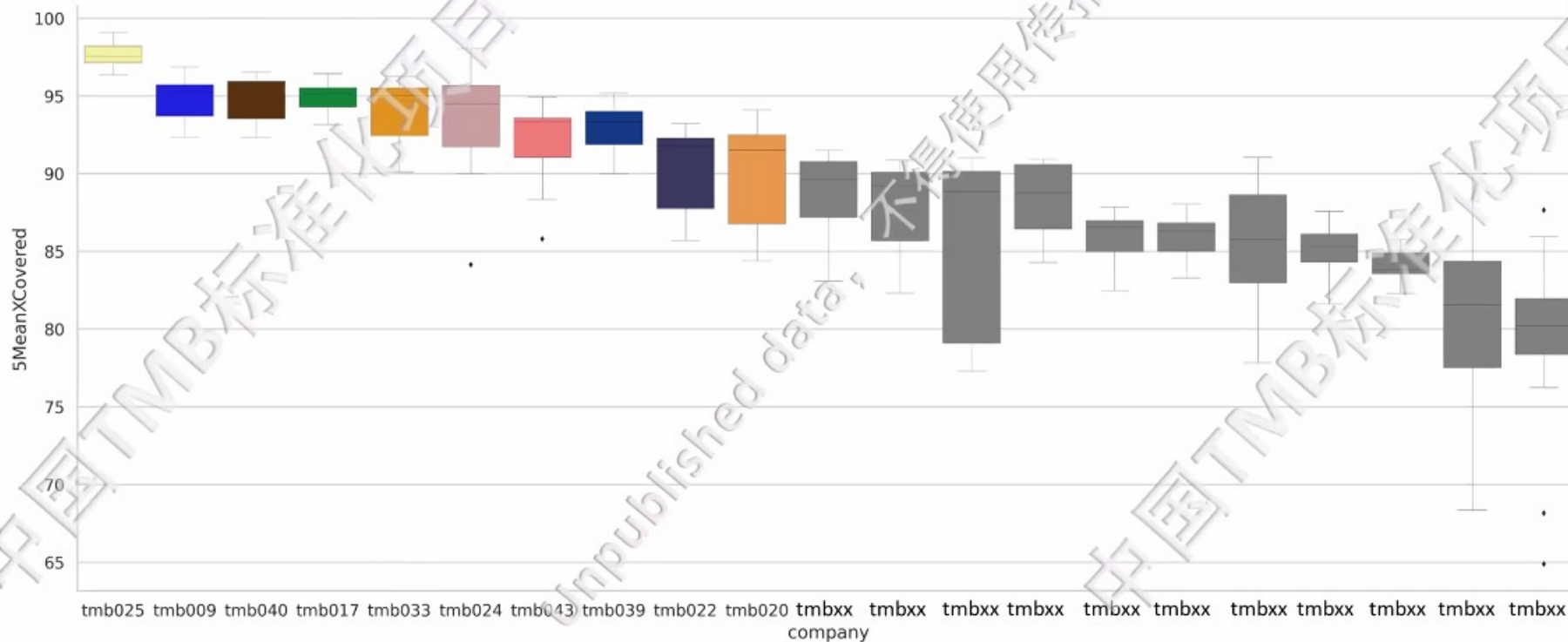
Q30则表示质量值≥30的碱基所占百分比，参与单位Q30数据均表现良好，所有参与单位均≥85%

参与单位质控数据——捕获效率



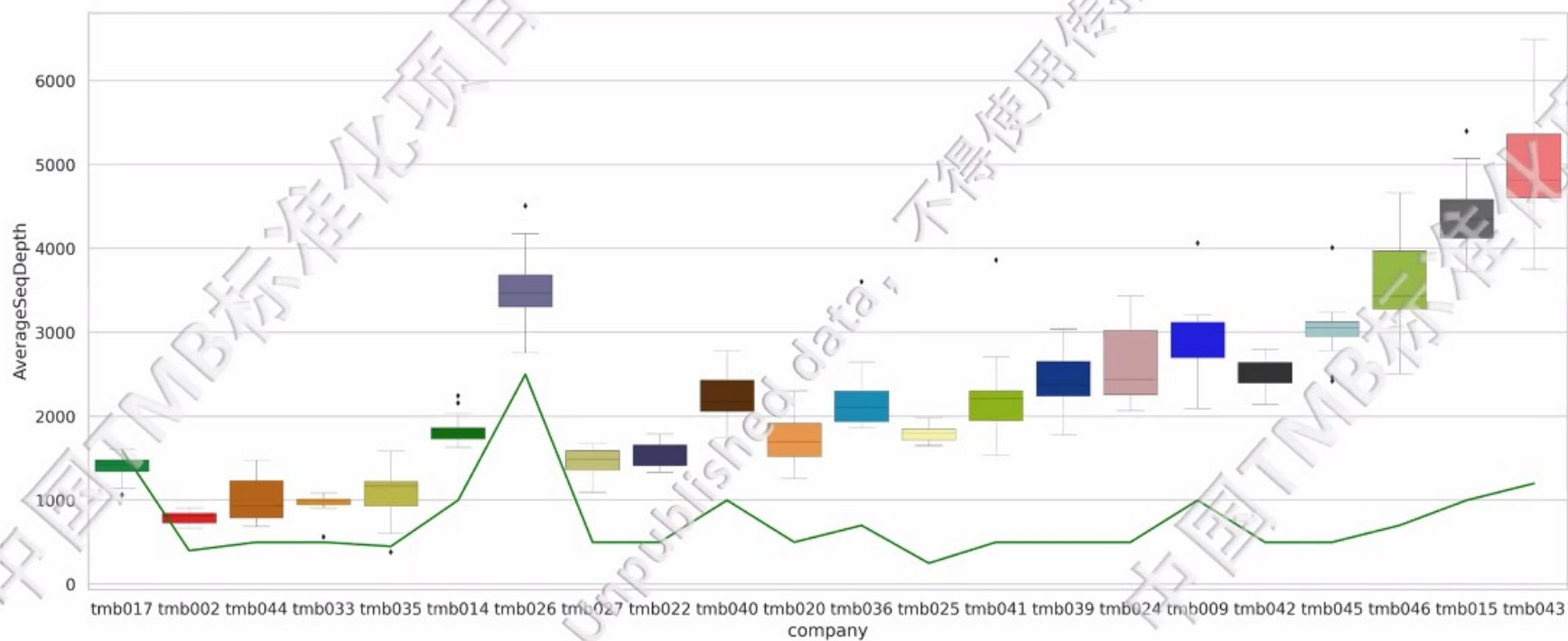
捕获效率是指测得比对到目标区域的碱基占全部比对到参考基因组的碱基的比例

参与单位质控数据——0.5X mean Coverd



0.5X mean Coverd: 至少覆盖平均深度*0.5的区域占目标区域的比例

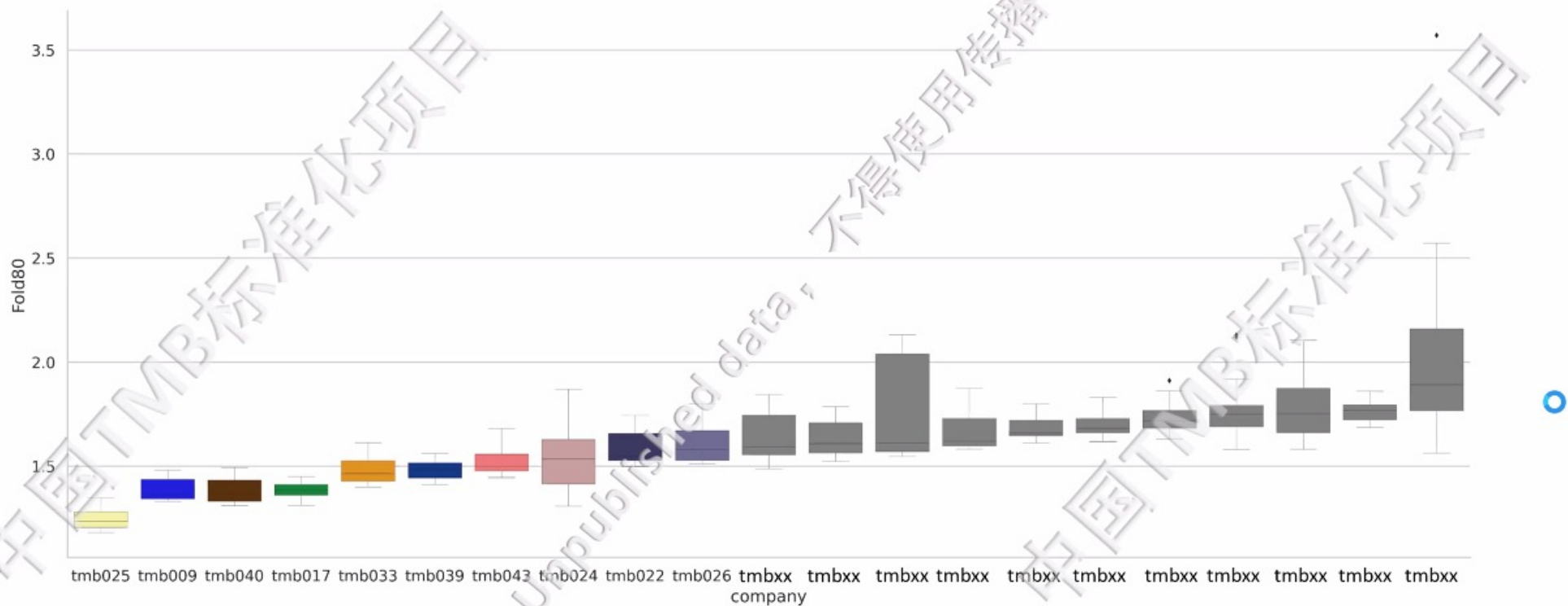
参与单位质控数据——去重后平均测序深度



去重后平均测序深度指可信的去重重复序列 (dup) 后目标区域的平均测序深度

折线为参与单位上报的去重后测序深度，箱图为标品结果去重后测序深度；

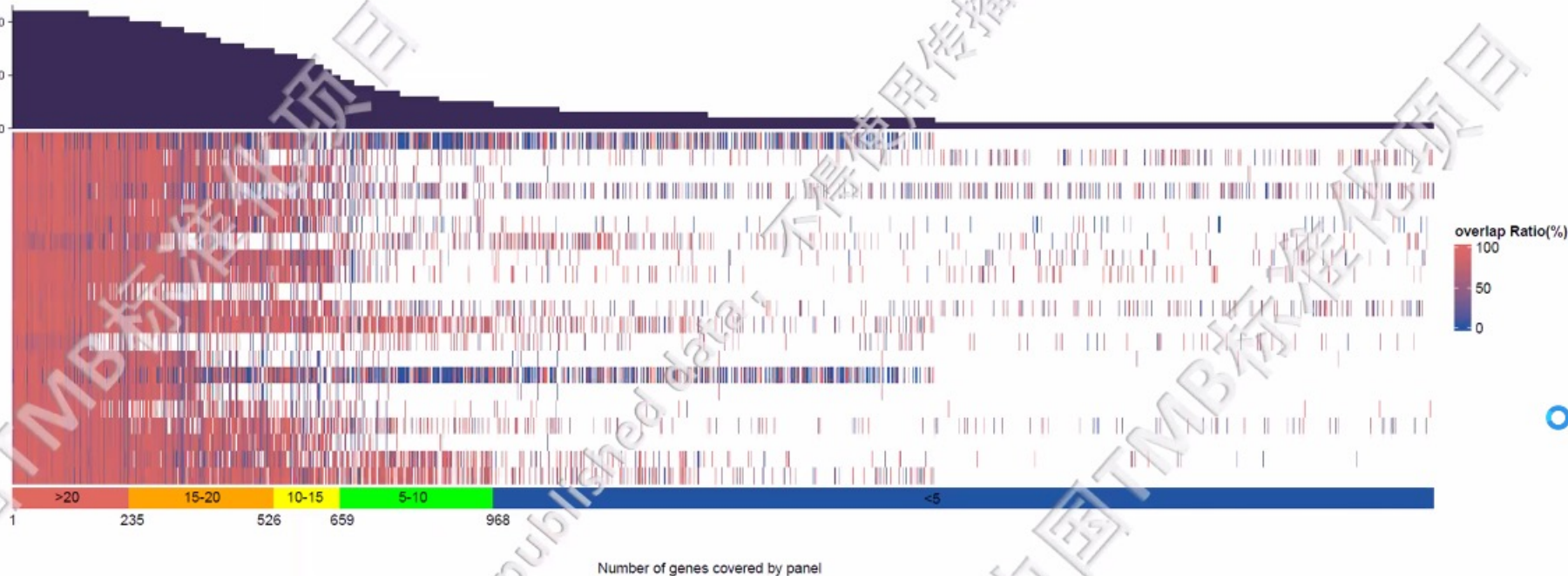
参与单位质控数据——均一性Fold80



均一性可以通过Fold80指标直观体现，Fold80计算公式为平均深度/(80%以上区域被覆盖的深度)，Fold80越低均一性越好。

各参与单位panel基本覆盖高频肿瘤突变基因，相似度较高

每个基因被
参与单位覆
盖的次数



参与公司提供的各家Panel共覆盖了2911个基因，其中200多个为高频热点基因。其余为仅少部分公司涉及或者各家公司特有的基因。该图说明在一些高频热点基因参与公司设计的试剂盒均有覆盖，但是又根据各家的产品特点，也独立设计了很多自己特有的检测区间。来满足临床检测治疗上的不同需求。

◆ 分析所用参考数据源：

<https://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/refGene.txt.gz>

统计各家参与单位Panel有不同的设计理念、方向

Panel的设计观察到三种情况：

➤ Group 1

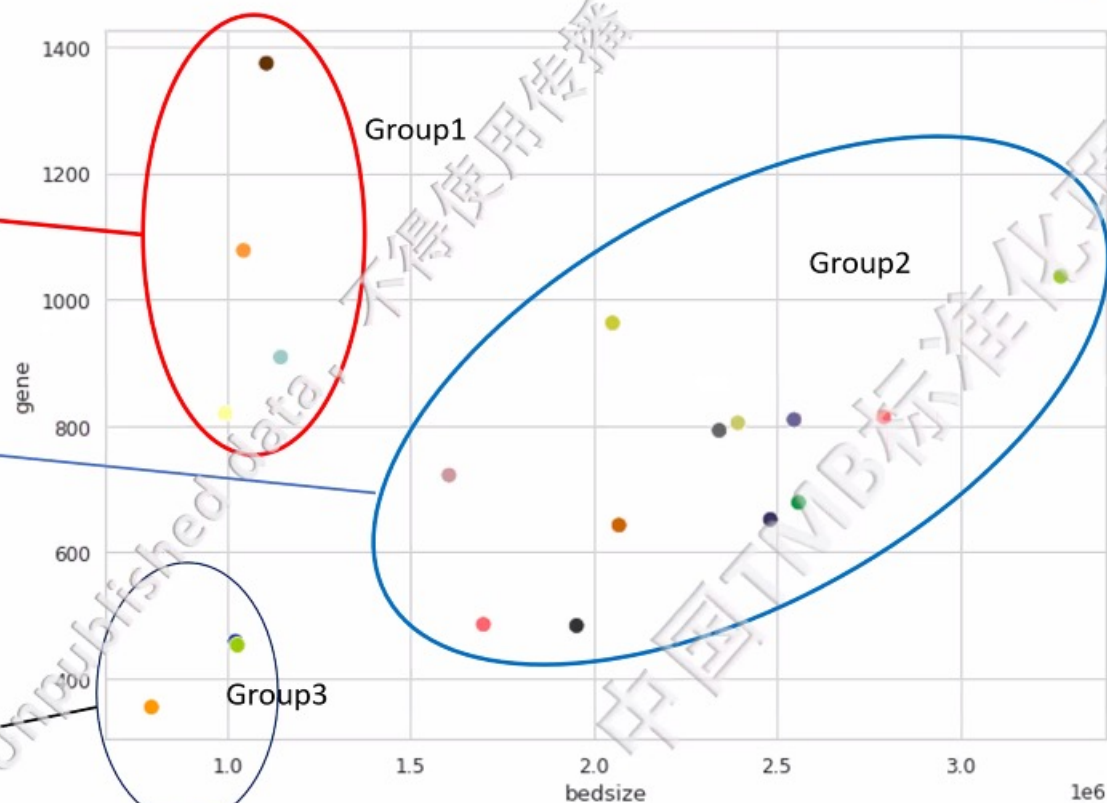
参与单位Panel基因数量/Panel长度 平均
>1000 gene/Mb, 设计策略为Panel区域较小,
基因数量较多

➤ Group 2

参与单位Panel基因数量/Panel长度平均为334
gene/Mb, 设计策略为Panel区域与基因数量
均较大

➤ Group 3

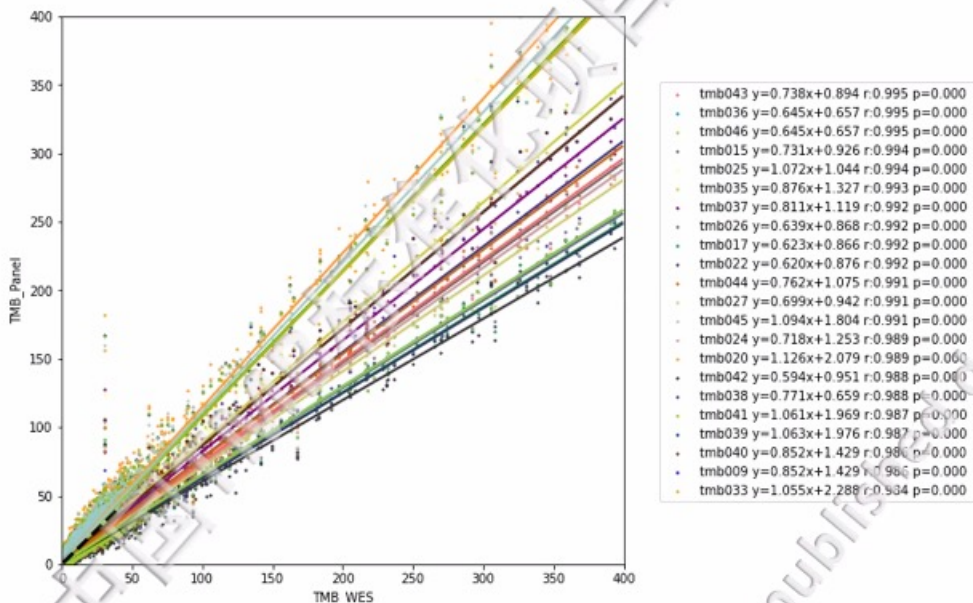
参与单位Panel基因数量/Panel长度平均为
446gene/Mb, 设计策略为Panel区域在1Mb左
右与基因数量较少



横坐标为Panel大小（单位Mb），纵坐标是覆盖的基因个数

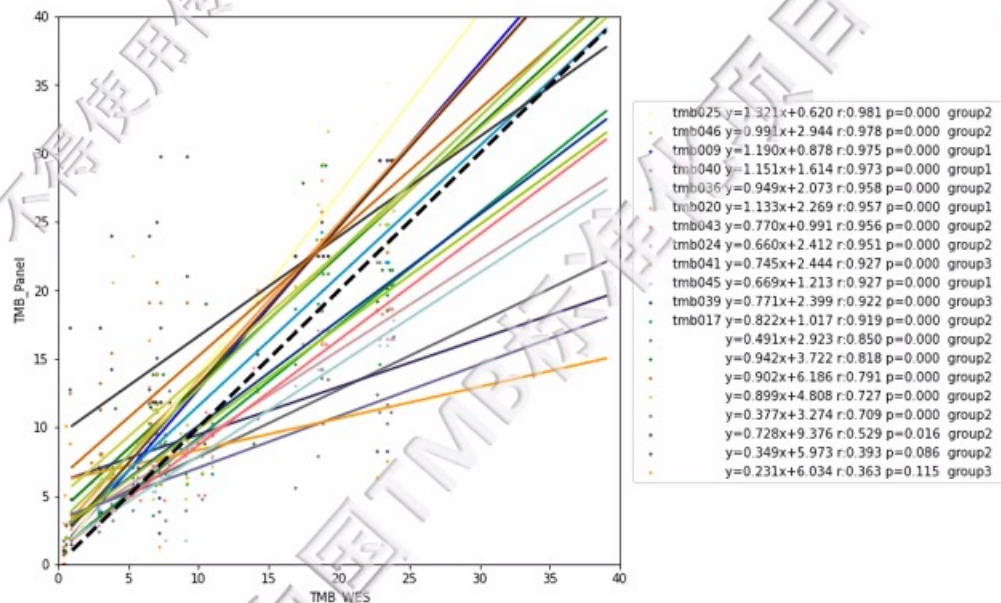
Panel-TMB与WES-TMB进行相关性评估

20例肿瘤参考品Panel-TMB与WES-TMB答案值一致性评估



TCGA数据集

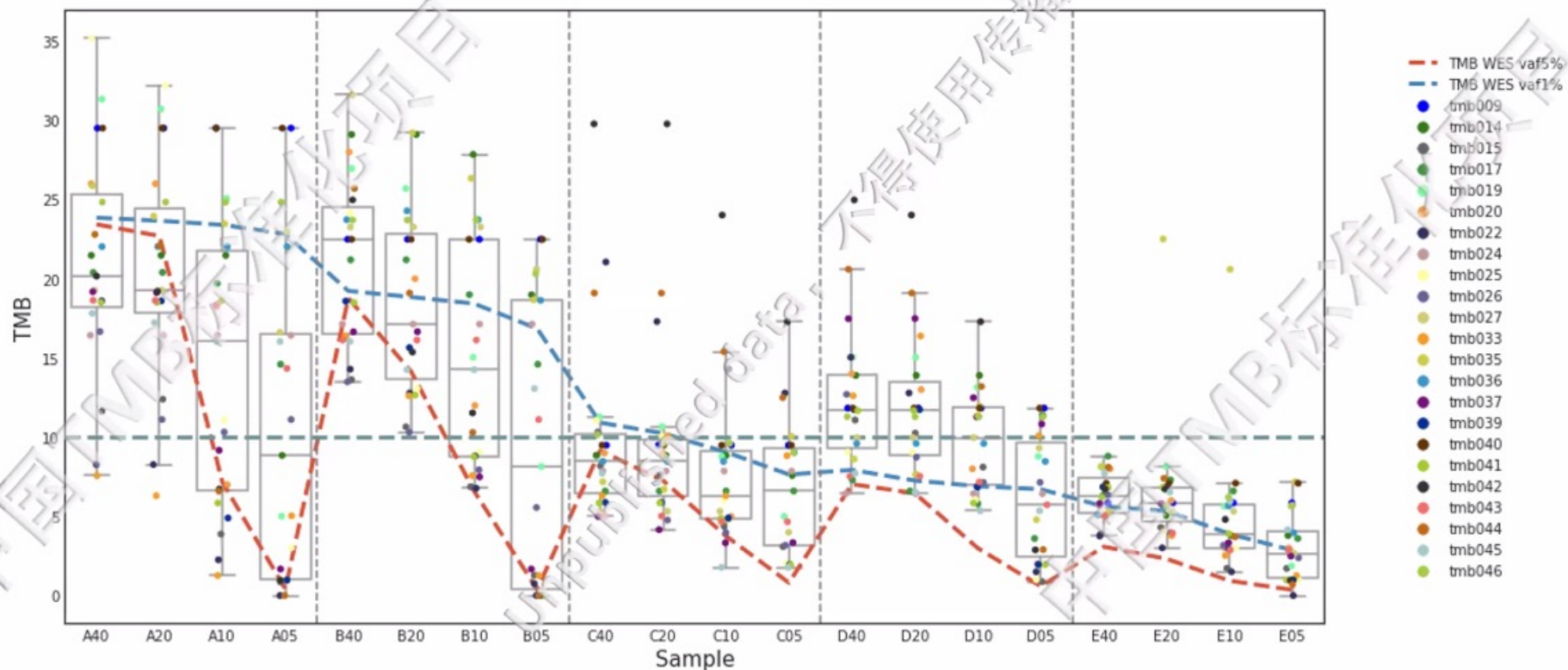
采用TCGA MC3数据，选取 $\geq 5\%$ 的突变位点，横坐标为四家公司全外交集bed计算WES-TMB，纵坐标为参与单位bed计算Panel-TMB，计算相关性



20例肿瘤参考品

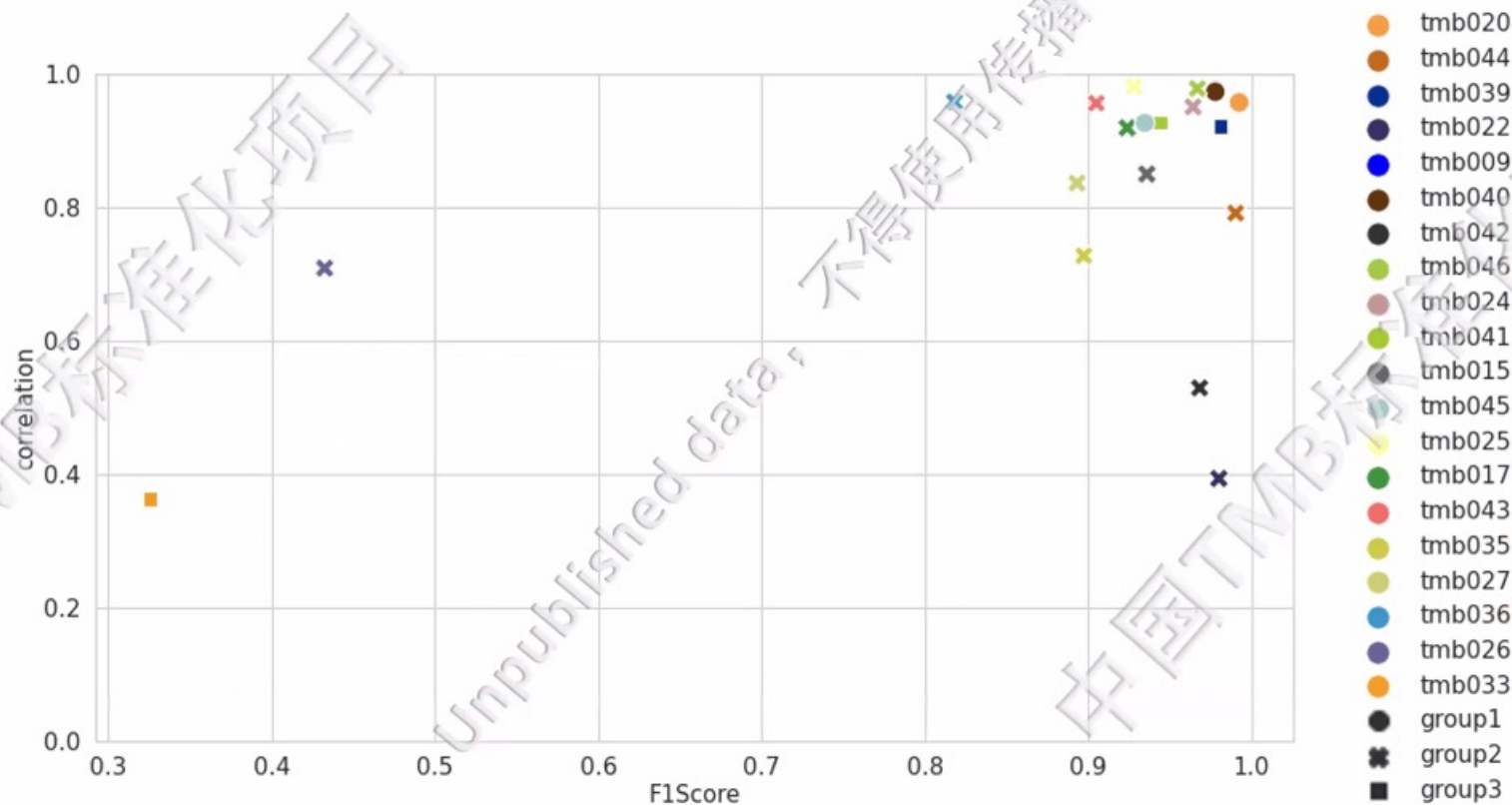
选取20例参考品，横坐标为各参与单位上报的计算TMB VAF计算WES-TMB值，纵坐标为各参与单位上报的Panel-TMB

不同肿瘤比例的参考品采用不同的WES TMB答案值进行分析



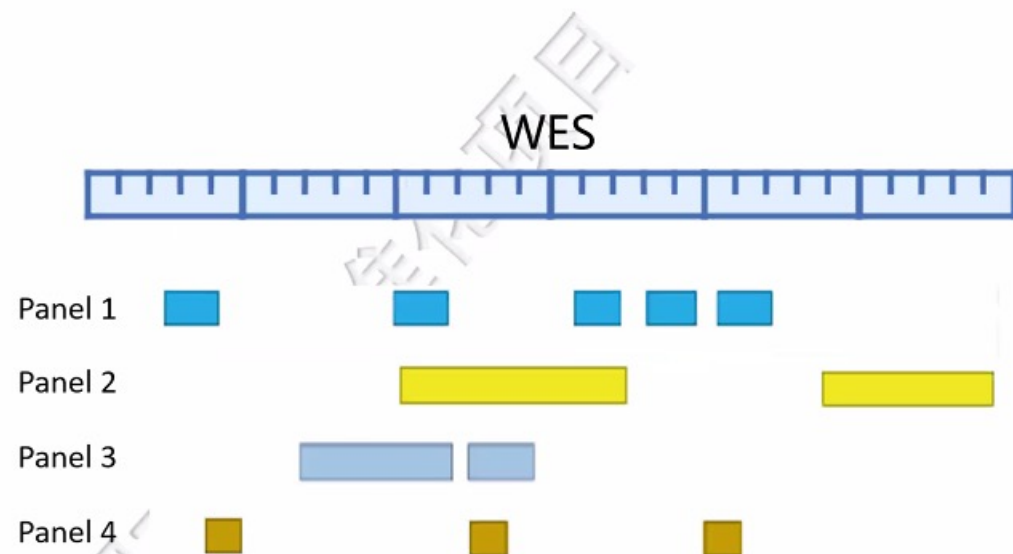
折线分别为按照5%和1%的VAF分别计算的答案集TMB值

参与单位位点准确度与TMB correlation之间关系



横坐标为按照参与单位提交计算TMB的VAF计算F1 score, 纵坐标为各参与单位Panel-TMB与WES-TMB的相关性

建立NGS测序panel标准化评估体系

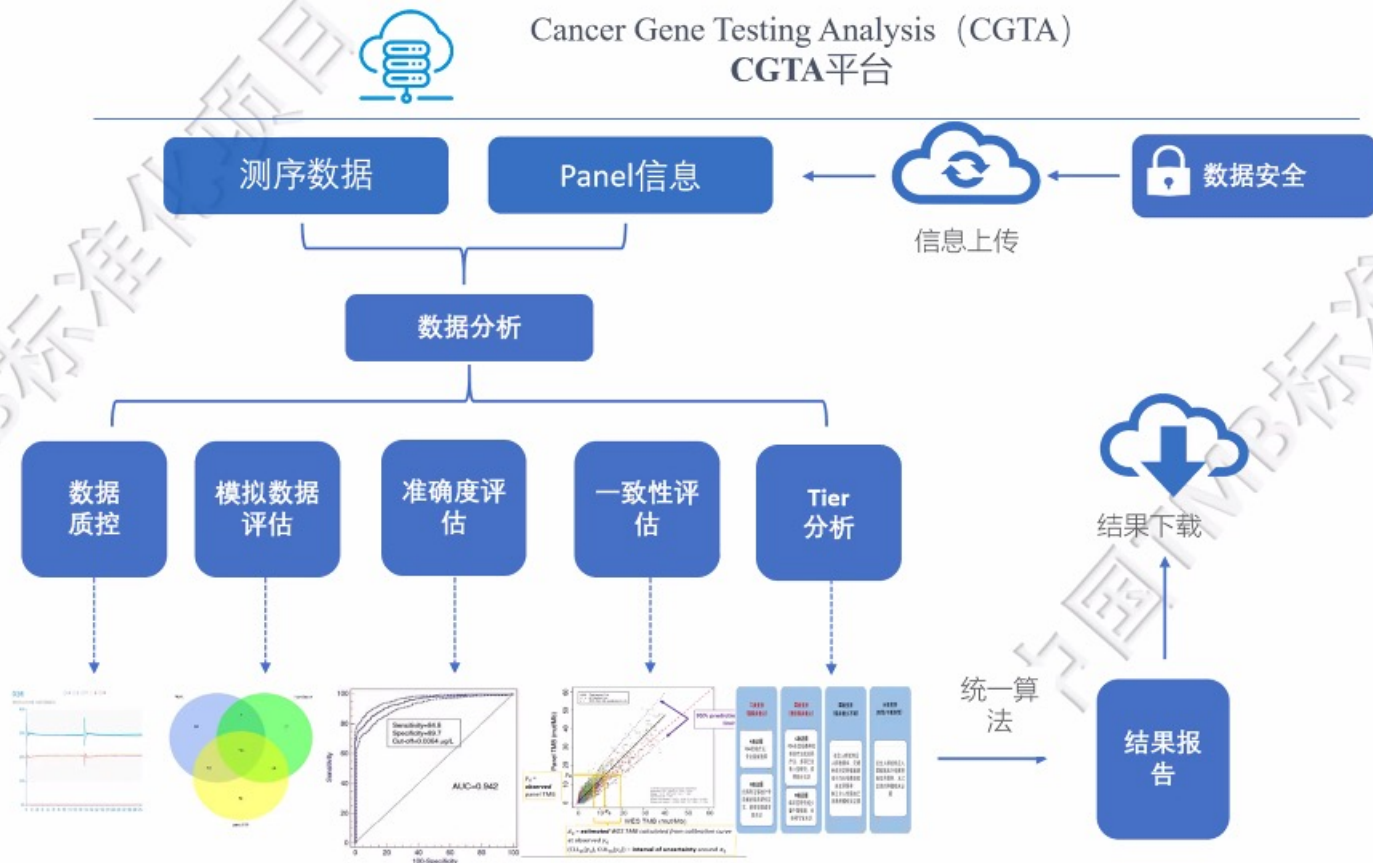


建立以WES数据为参照的数据库



建立靶向测序panel标准化评估体系

在项目一阶段结束后建立肿瘤大panel标准化评估网站





共同建立中国TMB标准!