

## CASE STUDY · IMMUNO-ONCOLOGY

# Finding, and confirming, markers of response to checkpoint blockade

How Biomarker Xplorer combines whole-genome, whole-exome and bulk RNA sequencing for discovery with targeted quantification for confirmation, in an illustrative immuno-oncology study.

**Illustrative case study.** This scenario shows how Biomarker Xplorer is applied to an immuno-oncology question. It does not describe a specific client project: cohort sizes, results and timelines are indicative and depend on indication, sample quality and study objectives. Markers are drawn from published immuno-oncology literature; their combination here is hypothetical and is not a validated clinical test.

## 1 Design

Groups, endpoints and success criteria fixed before any sample is run

## 2 Discover

WGS, WES and bulk RNA-seq on a balanced discovery cohort

## 3 Narrow

AI models reduce thousands of features to a testable few

## 4 Confirm

Targeted quantification on the full sample set, model locked

## At a glance

|                 |                                                                                                |
|-----------------|------------------------------------------------------------------------------------------------|
| QUESTION        | What separates patients who gain durable benefit from anti-PD-1 therapy from those who do not? |
| SETTING         | Clinical-stage biotech · phase II trial · anti-PD-1-based combination · advanced NSCLC         |
| GROUPS          | Durable clinical benefit (DCB) vs no durable benefit (NDB)                                     |
| FULL SAMPLE SET | 180 patients: pre-treatment tumour biopsy and baseline blood                                   |
| DISCOVERY       | 60 patients (30 DCB / 30 NDB): WGS, WES, bulk RNA-seq                                          |
| NARROWING       | ≈21,400 features → 11 testable candidates (AI models, nested cross-validation)                 |
| CONFIRMATION    | Targeted RNA panel (RT-qPCR), dPCR and amplicon DNA panel on 174 evaluable FFPE samples        |
| OUTCOME         | 7 confirmed markers in one locked score; held-out AUC 0.78 vs 0.62 for PD-L1 IHC alone         |
| DURATION        | ≈ 6 months                                                                                     |

## The challenge

PD-L1 immunohistochemistry and tumour mutational burden (TMB) are the most widely used predictors of benefit from immune checkpoint blockade, yet neither cleanly separates patients who respond from those who do not. A clinical-stage biotech running a phase II trial of an anti-PD-1-based combination in advanced non-small cell lung cancer (NSCLC) wanted to know which tumour features distinguish patients who gain durable benefit, to build an enrichment hypothesis for its next trial.

The material was good: a pre-treatment tumour biopsy and baseline blood for 180 patients. The risks were familiar:

- **Too many candidates, too few patients.** Genome-wide data yield tens of thousands of features for a few dozen patients per group, ideal conditions for false discoveries.
- **Markers that don't travel.** Signals found by sequencing frozen research cores often fail to transfer to the FFPE blocks and targeted assays a later trial would rely on.

**The brief:** find the markers that separate the two groups, and prove they hold up before building on them.

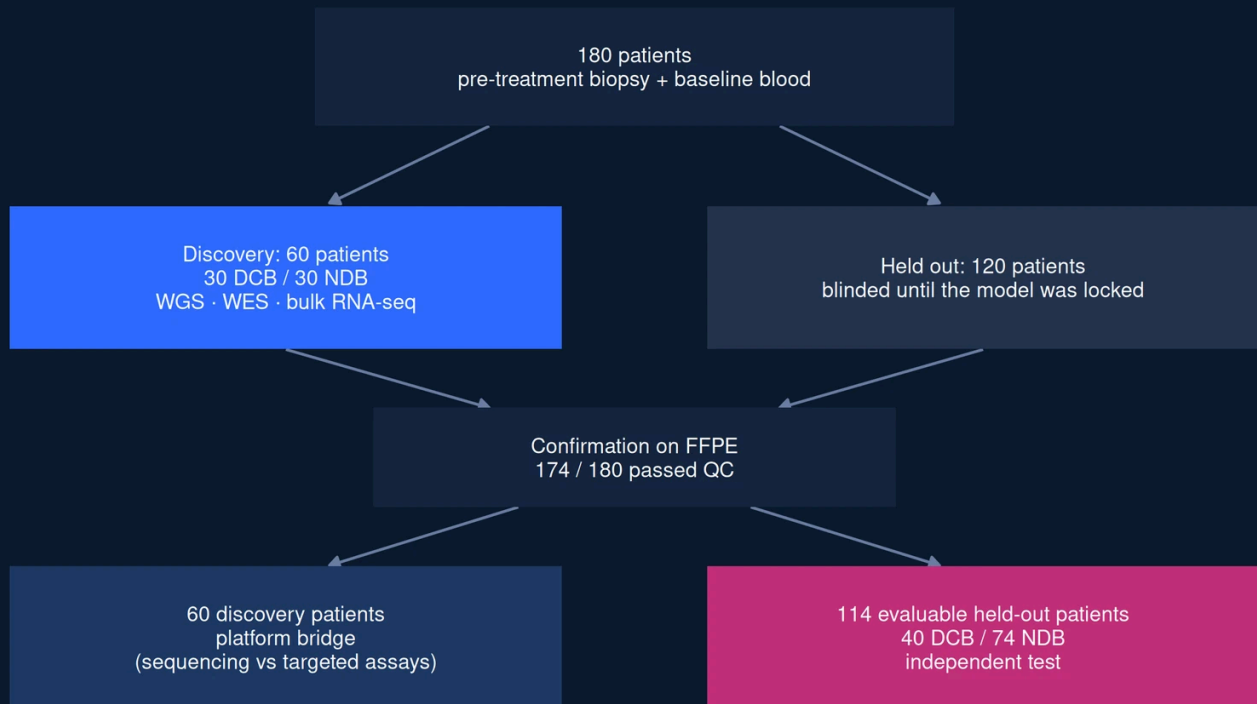
## 01 Design the study before running a sample

Most biomarker studies are decided at the design stage. Before any sample was processed, we agreed with the sponsor on:

- **Two groups, defined upfront.** Durable clinical benefit (DCB: complete or partial response, or stable disease  $\geq 6$  months, RECIST 1.1) versus no durable benefit (NDB: progression, or stable disease  $< 6$  months), the definition used in landmark NSCLC checkpoint studies<sup>1</sup>.
- **A discovery/confirmation split fixed in advance.** 60 patients with a frozen research core of sufficient quality formed the discovery cohort (30 DCB / 30 NDB), balanced for histology, PD-L1 tumour proportion score (TPS), smoking history and line of therapy. The other 120 patients were held out.
- **Blinding and batch randomisation.** Laboratory and analysis teams were blinded to outcome, and samples were randomised across extraction and sequencing batches so that batch effects could not masquerade as biology.
- **A pre-specified success criterion.** In held-out patients, the final model had to reach an AUC with a lower 95% confidence bound above 0.65 and add significant information to PD-L1 TPS, in line with PROBE principles for biomarker evaluation<sup>2</sup>.
- **Sample QC gates.** Pathology review of tumour cellularity, and DNA/RNA integrity thresholds (DIN, RIN, DV200) set for each assay.

## Study design: one sample set, two jobs

Split fixed before any sample was run



Illustrative design. 6 held-out samples had insufficient tumour content.

**Figure 1. Study design.** The 180 patients were split before any sample was run. The 60 with good-quality frozen tissue formed the discovery cohort (30 with durable benefit, DCB; 30 without, NDB); the other 120 were held out and kept blind. All FFPE blocks then went to targeted confirmation, and 174 passed QC. The 60 discovery patients bridge sequencing and targeted assays; the 114 evaluable held-out patients (40 DCB, 74 NDB) form the independent test. Illustrative design.

## 02 Profile the discovery cohort: three complementary layers

Each sequencing layer answers a different part of the immuno-oncology question. All three were run on the same 60 tumours, with matched blood as the germline reference.

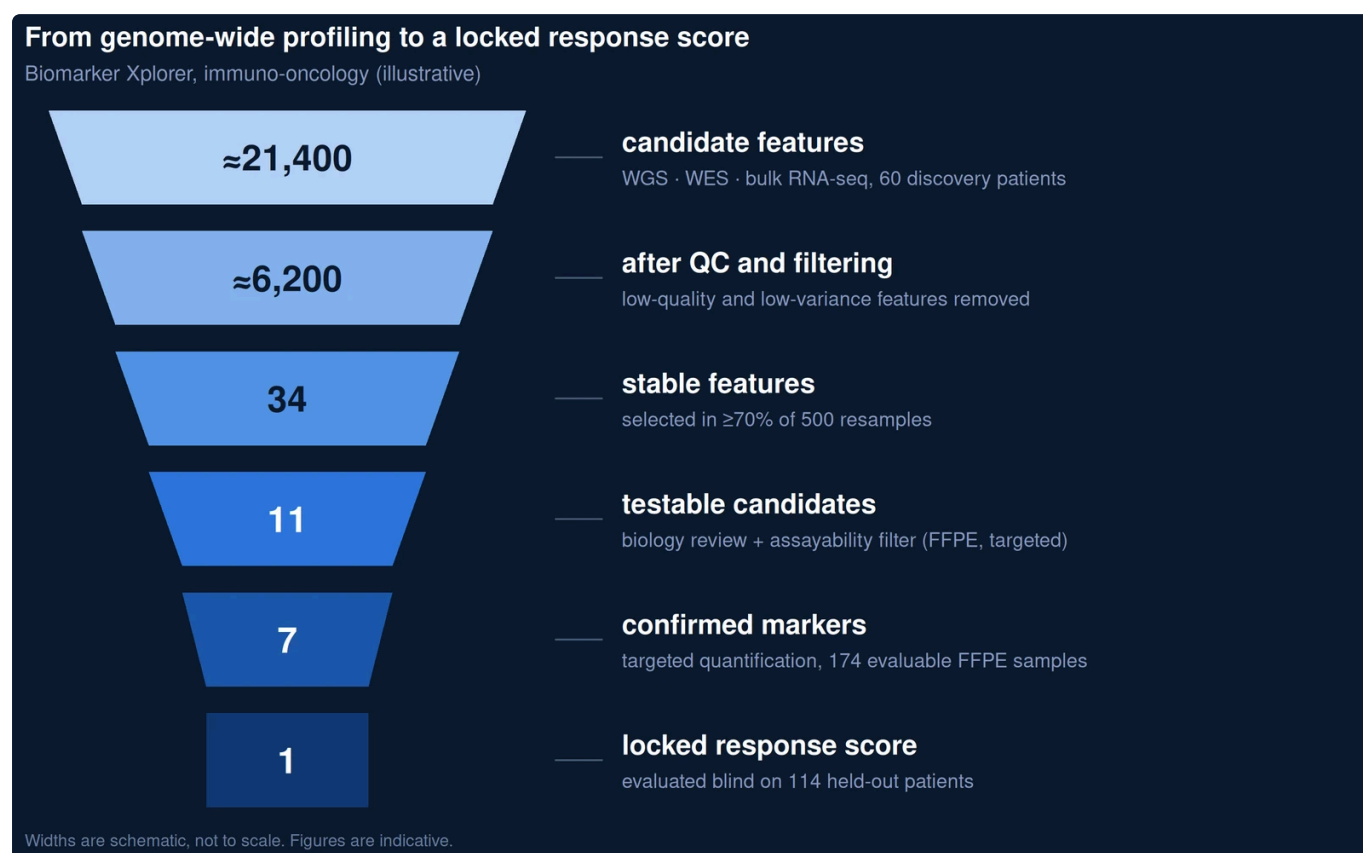
TABLE 1. THREE DISCOVERY LAYERS, ONE QUESTION

| Layer                         | Design                                                                    | What it contributes                                                                                                                                                                                                                        |
|-------------------------------|---------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Whole-genome sequencing (WGS) | Tumour $\approx 60\times$ / blood $\approx 30\times$ , PCR-free libraries | Genome-wide copy number and aneuploidy <sup>5</sup> , structural variants, whole-genome doubling and mutational signatures (tobacco, APOBEC, mismatch repair): events an exome captures poorly                                             |
| Whole-exome sequencing (WES)  | Tumour $\approx 200\times$ / blood $\approx 100\times$                    | Deep coding-variant calls: TMB and clonal TMB <sup>3</sup> , HLA class I typing and allele-specific HLA loss <sup>4</sup> , predicted neoantigens, and loss-of-function in immune-evasion and resistance genes (B2M, JAK1/2, STK11, KEAP1) |
| Bulk RNA-seq                  | Stranded total RNA, $\approx 50$ M read pairs                             | Immune programmes (IFN- $\gamma$ / T-cell-inflamed, cytotoxicity, B-cell / TLS), stromal TGF- $\beta$ response, antigen presentation, immune-cell deconvolution, expressed neoantigens and fusions                                         |

**Why both genome and exome?** Depth and breadth answer different questions. The exome's depth resolves the subclonal mutations needed for clonal TMB and neoantigen calls; the genome's breadth captures copy-number and structural events, the kind of alteration behind the 9p21 finding below.

## 03 Narrow thousands of candidates to a testable few

Profiling produced  $\approx 21,400$  features per patient: gene-expression values and signature scores, mutation and copy-number features, HLA and neoantigen metrics. The goal was not the best-looking model in discovery, but a short list of candidates likely to hold up in new patients.



**Figure 2. From genome-wide profiling to a locked score.** Each level gives the number of features still in play: about 21,400 from WGS, WES and RNA-seq; about 6,200 after quality filtering; 34 selected in at least 70% of 500 resamples; 11 that can be measured by a targeted assay on FFPE; 7 confirmed in held-out patients and combined into one locked score. Widths are schematic. Illustrative figures.

- **Leakage-proof modelling.** Filtering, normalisation and tuning were performed inside nested cross-validation (outer 5-fold  $\times$  20 repeats), the main safeguard against optimistic error estimates<sup>6</sup>.
- **Complementary models.** Penalised (elastic-net) logistic regression served as the interpretable primary model; gradient-boosted trees checked for non-linear effects and interactions.
- **Stability selection.** Only features selected in  $\geq 70\%$  of 500 resamples were kept<sup>7</sup>, 34 in total. SHAP explanations confirmed that each acted in a biologically plausible direction.
- **Assayability filter.** Every candidate had to be measurable on FFPE with a targeted assay at realistic tumour content. TMB, for example, was associated with benefit but added little once expression features were in the model, and reliable TMB needs a capture panel of  $\geq 1$  Mb<sup>8</sup>; it was kept as a covariate rather than forced into an amplicon proxy.

The result: 11 testable candidates, each mapped to a confirmation assay. Cross-validated AUC in discovery was 0.84 (95% CI 0.74–0.94), treated as optimistic until confirmed.

TABLE 2. THE 11 CANDIDATES TAKEN TO CONFIRMATION

| Candidate                                     | Biological rationale                                          | Higher in | Confirmation assay           |
|-----------------------------------------------|---------------------------------------------------------------|-----------|------------------------------|
| CXCL9, GZMB, IFNG                             | IFN- $\gamma$ / cytotoxic T-cell axis <sup>9</sup>            | DCB       | Targeted RNA panel (RT-qPCR) |
| CD274 (PD-L1 mRNA)                            | Adaptive immune resistance                                    | DCB       | Targeted RNA panel (RT-qPCR) |
| CXCL13, MS4A1                                 | B-cell / tertiary lymphoid structure axis <sup>10</sup>       | DCB       | Targeted RNA panel (RT-qPCR) |
| TGFBI, ACTA2                                  | Stromal TGF- $\beta$ response, T-cell exclusion <sup>11</sup> | NDB       | Targeted RNA panel (RT-qPCR) |
| STK11, KEAP1 loss-of-function                 | Immune-cold lung adenocarcinoma <sup>12</sup>                 | NDB       | Amplicon DNA panel           |
| 9p21.3 deletion (CDKN2A / type I IFN cluster) | Cold microenvironment, primary resistance <sup>13</sup>       | NDB       | dPCR copy-number assay       |

#### 04 Confirm by targeted quantification on the full sample set

Three targeted assays were designed and analytically checked, efficiency, linearity, limit of detection and repeatability, following MIQE and dMIQE guidance<sup>14,15</sup>, then run on FFPE sections from all 180 patients. 174 passed QC (6 held-out samples had insufficient tumour content). The model was locked before outcomes for held-out patients were unblinded.

- **Targeted RNA panel (RT-qPCR)**: 8 targets plus 3 reference genes, robust to fragmented FFPE RNA.
- **Digital PCR (dPCR)**: absolute 9p21.3 / reference-locus copy-number ratio, corrected for tumour cellularity.
- **Amplicon DNA panel**: STK11 and KEAP1 at  $\geq 500\times$  coverage, with B2M, JAK1 and JAK2 reported descriptively (too rare in this cohort to model).

The full sample set did two jobs. The 60 discovery patients bridged platforms, frozen-tissue sequencing versus FFPE targeted assays, and the 114 evaluable held-out patients provided the independent test.

TABLE 3. PLATFORM BRIDGE IN THE 60 DISCOVERY PATIENTS

| Readout                        | Comparison            | Agreement                          |
|--------------------------------|-----------------------|------------------------------------|
| Gene expression                | RT-qPCR vs RNA-seq    | Spearman $\rho$ 0.81–0.93 per gene |
| Mutation status (STK11, KEAP1) | Amplicon panel vs WES | 58 / 60 patients concordant        |
| 9p21.3 copy number             | dPCR vs WGS           | Cohen's $\kappa$ = 0.86            |

*Frozen cores and FFPE blocks are different pieces of tumour; some discordance reflects intratumour heterogeneity, not assay error.*

TABLE 4. OUTCOME OF THE 11 CANDIDATES IN HELD-OUT PATIENTS

| Outcome                  | Markers                                                       | Interpretation                                                                                                       |
|--------------------------|---------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------|
| <b>Confirmed (7)</b>     | CXCL9, GZMB, CXCL13, CD274, TGFBI, STK11 LoF, 9p21.3 deletion | Significant in the pre-specified direction                                                                           |
| <b>Redundant (2)</b>     | IFNG, ACTA2                                                   | IFNG: low abundance in FFPE, no information beyond CXCL9.<br>ACTA2: captured by TGFBI.                               |
| <b>Not confirmed (2)</b> | MS4A1, KEAP1 LoF                                              | MS4A1: signal explained by CXCL13.<br>KEAP1: same direction, but the 95% CI included no effect: kept as exploratory. |

TABLE 5. LOCKED SCORE IN HELD-OUT PATIENTS (N = 114; 40 DCB / 74 NDB)

| Model                                        | AUC (95% CI)            | Comment                                                      |
|----------------------------------------------|-------------------------|--------------------------------------------------------------|
| <b>Xplorer IO response score (7 markers)</b> | <b>0.78 (0.69–0.87)</b> | Pre-specified criterion met                                  |
| <b>PD-L1 IHC (TPS) alone</b>                 | 0.62 (0.51–0.73)        | Reference                                                    |
| <b>Score + TPS vs TPS alone</b>              | –                       | Significant added value (likelihood-ratio test, $p < 0.01$ ) |

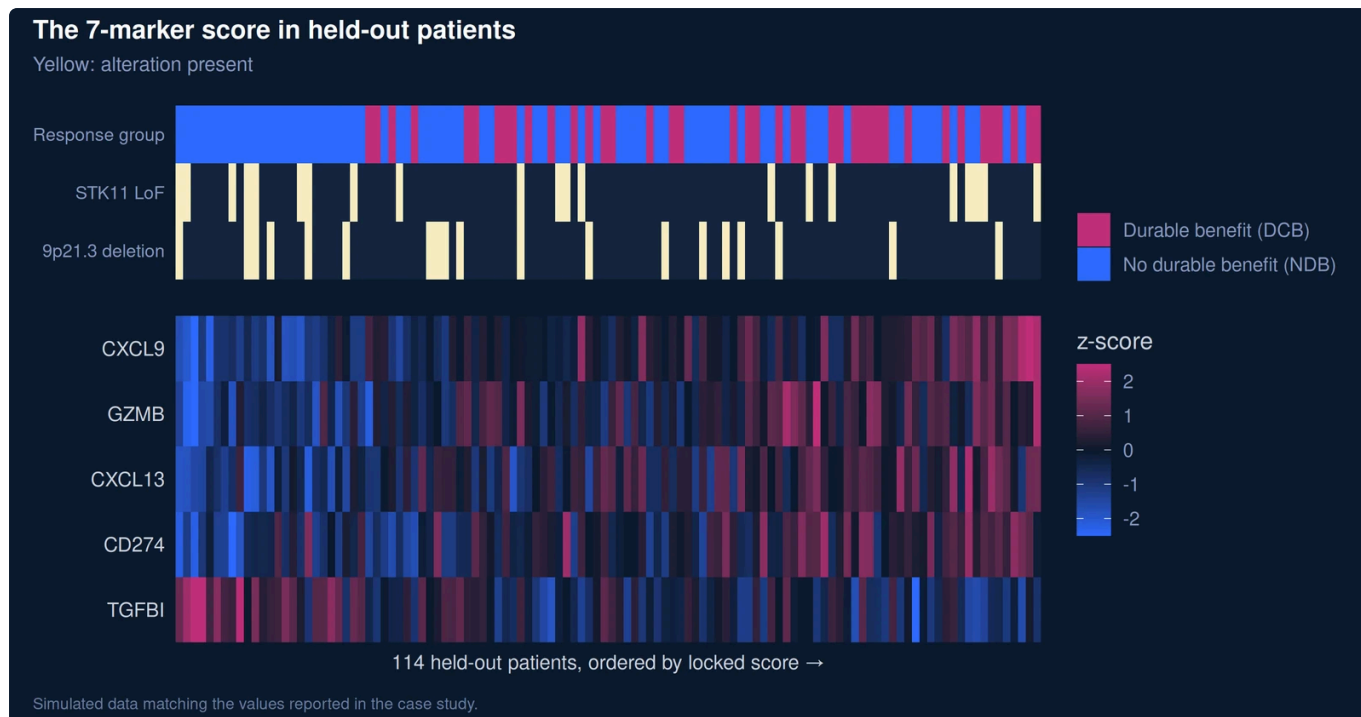
The drop from 0.84 in discovery to 0.78 in held-out patients is expected, and it is the held-out number that a development programme can build on.

### Fate of the 11 candidates in held-out patients

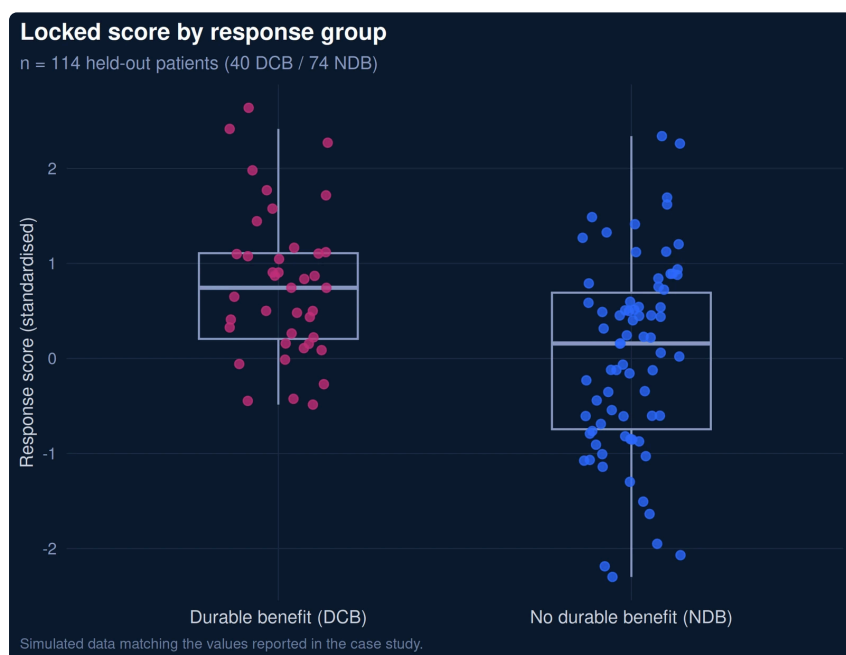
Arrow: group in which the marker is higher



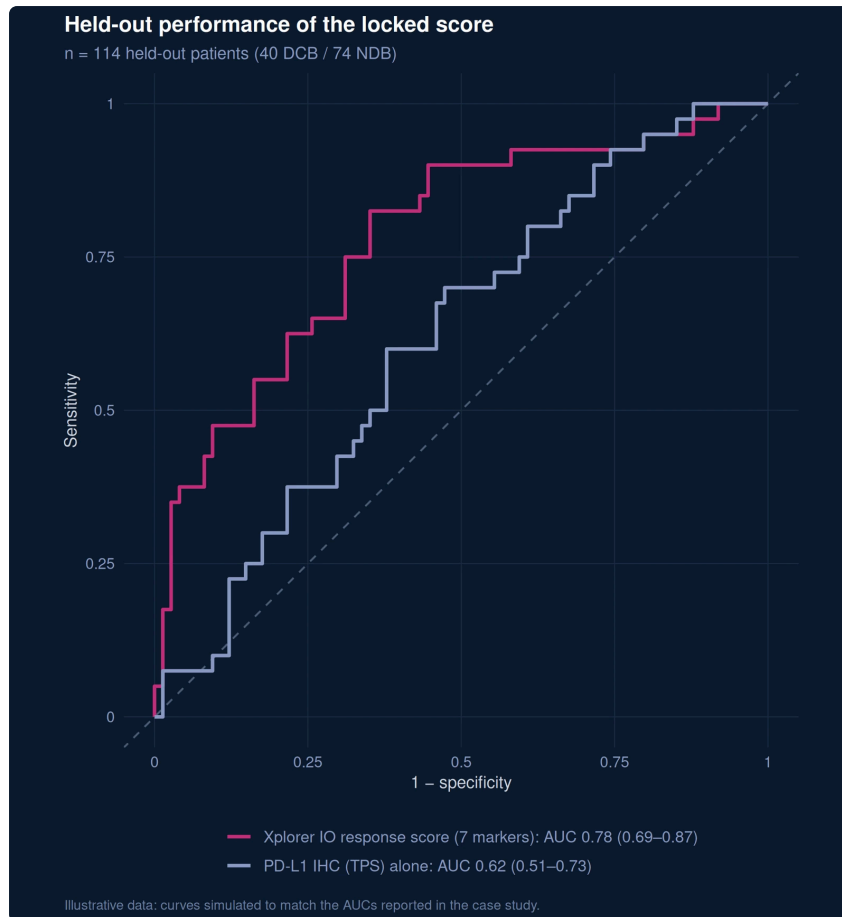
**Figure 3. Fate of the 11 candidates in held-out patients.** Each tile is one candidate; the arrow gives the group in which it is higher (DCB: durable benefit, NDB: no durable benefit). Seven were confirmed in the pre-specified direction. Two were redundant: IFNG added nothing beyond CXCL9, and ACTA2 was captured by TGFBI. Two were not confirmed: MS4A1 was explained by CXCL13, and KEAP1 pointed the same way but its 95% CI included no effect. Illustrative results.



**Figure 4. The 7-marker score in held-out patients.** Each column is one of the 114 held-out patients, ordered from lowest to highest score. Top: response group, and presence (yellow) of STK11 loss-of-function or 9p21.3 deletion, two markers more frequent in patients without durable benefit. Bottom: expression of the five genes of the score, as z-scores (pink above average, blue below). CXCL9, GZMB, CXCL13 and CD274 rise with the score; TGFBI falls. Simulated data matching the case study.



**Figure 5. Locked score by response group.** Each point is one held-out patient; boxes show the median and interquartile range. Scores are higher in patients with durable benefit, but the two groups overlap: the score ranks patients by likelihood of benefit, it does not classify every patient correctly. Simulated data matching the case study.



**Figure 6. ROC curves in held-out patients.** For every possible cut-off, a curve plots the share of DCB patients correctly identified (sensitivity) against the share of NDB patients wrongly flagged ( $1 - \text{specificity}$ ); the diagonal is chance. The 7-marker score (AUC 0.78) separates the groups better than PD-L1 IHC alone (AUC 0.62), and adding it to PD-L1 TPS gives significant added value (likelihood-ratio test,  $p < 0.01$ ). Curves simulated to match the reported AUCs.

## What the sponsor received

- A **locked 7-marker response score**, five-gene expression module, STK11 status and 9p21.3 copy number, with coefficients, cut-off and analysis code.
- **Three FFPE-compatible targeted assays** with SOPs and analytical-performance data, ready for formal analytical validation.
- The **integrated, quality-controlled WGS, WES and RNA-seq dataset** (FASTQ, BAM/CRAM, variant and expression matrices) and a full statistical report.
- A **documented no-go** on four candidates, and on TMB as a stand-alone selector in this setting, development effort not spent on markers that would not hold up.
- A **recommended next step**: prospective testing of the score as an enrichment strategy, with optional spatial and single-cell follow-up to locate the CXCL13 / TLS signal in tissue.

#### INDICATIVE TIMELINE ( $\approx$ 6 MONTHS)

|             |                                                                                |
|-------------|--------------------------------------------------------------------------------|
| Weeks 0–3   | Study design, statistical analysis plan, sample QC                             |
| Weeks 3–12  | Discovery sequencing (WGS, WES, RNA-seq) and bioinformatics QC                 |
| Weeks 10–16 | Data integration, AI modelling, candidate selection                            |
| Weeks 16–19 | Targeted assay design and analytical checks                                    |
| Weeks 19–25 | Confirmation run on the full sample set, locked-model evaluation, final report |

## Why it works

- **Decide before you measure.** Groups, endpoints and success criteria are fixed before data exist.
- **Discover broadly, confirm narrowly.** Genome-wide data to find candidates; targeted assays to confirm them on every sample.
- **Build assayability into selection.** Every candidate leaves discovery with a matching targeted assay.
- **Confirm blind, with a locked model.** The held-out result is the one you can build on.
- **Treat negative results as results.** Knowing which markers fail is what protects the next trial.

## Beyond this example

The same design applies across immuno-oncology questions:

- **Response and resistance** to checkpoint inhibitors, bispecifics, ADCs or cell therapies; primary versus acquired resistance using paired baseline and progression biopsies (e.g. JAK1/2 or B2M loss<sup>16</sup>).
- **Tumour immune phenotypes**, inflamed, excluded or desert, and their genomic drivers.
- **Toxicity:** immune-related adverse events, using germline and blood-based profiling.
- **Indications** such as melanoma, renal cell carcinoma, urothelial and head-and-neck cancers, and MSI-high colorectal cancer; sample types including fresh-frozen tissue, FFPE and blood.

## References

1. Rizvi NA, et al. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science*. 2015;348:124–128.
2. Pepe MS, et al. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. *J Natl Cancer Inst*. 2008;100:1432–1438.
3. McGranahan N, et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science*. 2016;351:1463–1469.
4. McGranahan N, et al. Allele-specific HLA loss and immune escape in lung cancer evolution. *Cell*. 2017;171:1259–1271.
5. Davoli T, et al. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science*. 2017;355:eaaf8399.
6. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*. 2006;7:91.
7. Meinshausen N, Bühlmann P. Stability selection. *J R Stat Soc Series B*. 2010;72:417–473.

8. Buchhalter I, et al. Size matters: dissecting key parameters for panel-based tumor mutational burden analysis. *Int J Cancer*. 2019;144:848–858.
9. Ayers M, et al. IFN- $\gamma$ -related mRNA profile predicts clinical response to PD-1 blockade. *J Clin Invest*. 2017;127:2930–2940.
10. Cabrita R, et al. Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature*. 2020;577:561–565.
11. Mariathasan S, et al. TGF $\beta$  attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature*. 2018;554:544–548.
12. Skoulidis F, et al. STK11/LKB1 mutations and PD-1 inhibitor resistance in KRAS-mutant lung adenocarcinoma. *Cancer Discov*. 2018;8:822–835.
13. Han G, et al. 9p21 loss confers a cold tumor immune microenvironment and primary resistance to immune checkpoint therapy. *Nat Commun*. 2021;12:5606.
14. Bustin SA, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem*. 2009;55:611–622.
15. dMIQE Group, Huggett JF. The digital MIQE guidelines update: minimum information for publication of quantitative digital PCR experiments for 2020. *Clin Chem*. 2020;66:1012–1029.
16. Zaretsky JM, et al. Mutations associated with acquired resistance to PD-1 blockade in melanoma. *N Engl J Med*. 2016;375:819–829.

### Planning a biomarker study?

Tell us your two groups, your samples and the decision you need to make. We will propose a design, discovery layers, cohort split and confirmation assays, before a single sample is run.

[genxmap.com/contact](https://genxmap.com/contact)