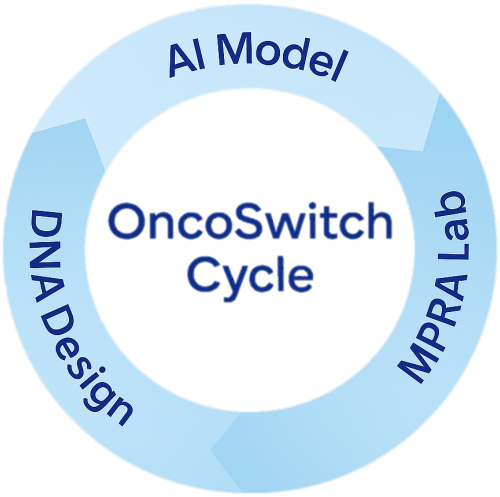
ONCOSWITCH

AI-programmable DNA switches for selective, safe, and adaptive gene therapy

Scientific and Commercial Roadmap

2025-2026



Angel round target: $1.2M

10 experimental cycles

18 months until the first licence

3-5 patentable DNA switches

https://www.oncoswitch.ai

malika@oncoswitch.ai
vitalii@oncoswitch.ai

INTRODUCTION

CONTEXT AND PHILOSOPHY OF THE PROJECT

Each of us knows someone who has been affected by cancer. For some, it is their parents, for others, their friends, and for others, it is their own story. We do not romanticise the disease or promise miracles.

We are building a tool that makes treatment smarter and safer: DNA switches that activate therapeutic genes in tumours and remain silent in normal tissue.

This is not a beautiful metaphor – it is engineering work, where the result is measured by clear metrics and by how much the risk has been reduced for a specific patient.

18 months
Before the first license

10,000+
Verified sequences

8 - 12 weeks
Cycle duration

AI ↔ Laboratory → Data

"We are creating the technological foundation for the medicines of the future."

We are people of action who care. We have chosen the path of consistent and productive work rather than loud promises: day after day, we bring the idea to a form that can be trusted. We are not building a one-off trick in a test tube, but a foundation — a platform that makes therapy smarter and kinder to the patient. A platform that can be scaled, passed on, and integrated into other processes — so that in a year's time, it will be standard practice in the treatment of such diseases. We want patients to have the

chance to undergo treatment without unnecessary pain. That is why we are here.

Our motivation is simple and transparent: we want effective treatment to cease being synonymous with dangerous and toxic. We want doctors to be able to fine-tune their approach, rather than just going full throttle. We want therapy developers to have a switch that allows them to tailor treatment according to the principle of “only here, only now”. We choose meaning and solutions over noise. We stand for truth and quality, not pretty promises. We stand for transparency and quality control, not dazzling presentations. We are people of action. First, we patent the sequences, and then we tell the world about them. We calmly accept the failures of individual candidates and turn them into experience that strengthens the platform and its data. We don't trade in slides – we open the platform where you can see the live work: how the model matures, how controls are maintained, and how every dollar of the cycle is spent.

Our goal:

«SMART GENE THERAPY WITHOUT PAIN OR TOXICITY»

We understand very well that behind every graph there is a person. That is why our materials contain more than just numbers and beautiful curves. We clearly specify the deadlines and quality of work in advance, give simple technical tasks for pilots, and openly describe the terms of licences for sequences. We honestly show where a ‘ladder’ of verification is needed: first, a cell line, then three-dimensional tissue models in a dish (organoids), then primary cells, and only then small animal studies. We do not hide negative results: they also move the system forward and make it safer.

Our promise is simple: each cycle is not just a new set of numbers. It is another step towards the day when cancer therapies will be customisable,

gentle, and as effective as possible. We are doing our part on this journey — carefully, stubbornly, and with respect for those for whom it all began.

SCIENTIFIC AND TECHNICAL VISION AND MISSION

VISION

We want the design of regulatory DNA sequences to become an industry standard: switches activate therapy in tumours and remain silent in normal tissue.

How it works in essence. We use artificial intelligence not as a black box, but as a ‘guiding hand’ for the experiment. In a conventional neural network, numbers serve as ‘weights.’ Instead of numbers, we use DNA letters. We set the model's goal as ‘high activity in tumours, low activity in normal tissue,’ and it suggests a set of candidate sequences. we test them in parallel (MPRA, Massive Parallel Reporter Assay – a test that shows the activity of given

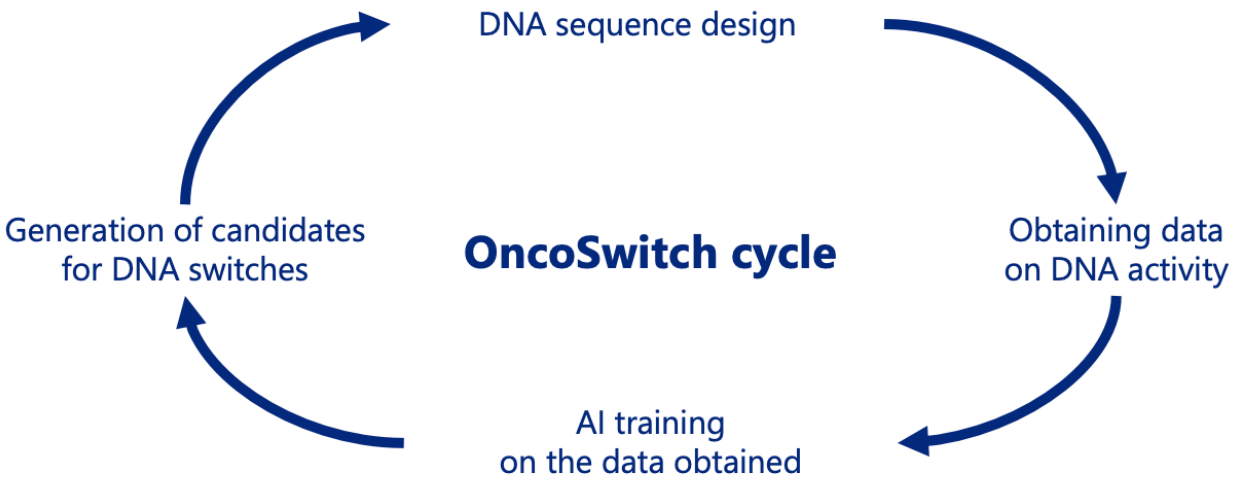
MISSION

To provide biotech and pharmaceutical companies with a working tool – a technological and laboratory platform with a closed cycle of design → testing (MPRA) → training → redesign. In a single run, we test thousands of DNA sequences in parallel; the activity measurements obtained in the ‘tumour/normal’ pair are fed back into the model. The model learns from real data, rethinks the rules and suggests the best options for the next iteration. This is how we are getting closer to truly selective switches, step by step.

Partners work through a web office and API: they see projects in real time, receive validated switches

DNA sequences in cells), return the actual measurements back to the model, and it reassembles the design. This creates a closed loop: design → test → training → redesign, until we obtain stable, safe switches for the selected context.

and reports, and then follow a clear path to licensing.



KEY DIFFERENCE.

We do not sell one-off designs, but a **scalable platform**.

- online service [simple at first, but with each experiment and each joint study, its capabilities and flexibility will grow]
- a closed dataset [similar to the service, it grows with each experiment],
- a learning model [becomes ‘smarter’ with each experiment, guides research more accurately, finds extraordinary solutions]
- IP on sequences [we patent the best working options].

Each MPRA cycle increases the variability and flexibility of our models, data, and platforms, improves quality metrics, accelerates and refines the production of new sequences by improving models, and expands the switch library.

PROJECT STRUCTURE AND PHASES

extended scientific and business version

Each phase of the project includes scientific experiments and engineering tasks, as well as strategic business outcomes. This section details the laboratory logic behind each stage, linked to investment and commercial dynamics.

PHASE 1. INITIALISATION OF THE PLATFORM CYCLE

0-60 DAYS

SCIENTIFIC DESCRIPTION

We select the first pair of "**tumor/normal**" considering biosafety and compatibility with MPRA (in particular, lenti-MPRA), as well as the possibility of further expansion into the CAR-T therapy market. We draw up and record a protocol (SOP, standard operating procedure) containing a clear algorithm for our laboratory and technological platform. Additionally, we **validate the laboratory circuit**: we approve the panel of control samples and barcode saturation, and establish **quality control threshold criteria** and the procedure for restarting stages. The result is a pipeline ready to launch the first cycle without any on-the-fly modifications.

We hire a team of key roles (discussed below) and draw up an agreement/contract with the service laboratory, which will be the base for the research cycles. The laboratory must be fully equipped to work with cellular and bioinformatic technologies (cell culture, MPRA method, sequencing).

At the same time, we begin to deploy the platform layout: graphic and user interface components, client-server interaction settings, platform security settings, basic visualizations and animations, and model version control.

We prepare the **starting library** for the first MPRA cycle: ~4000 sequences; active insert 50 bp (excluding technical additions: primer sites/linkers for cloning, barcodes/UMIs for tracking, service adapters). We include positive and negative controls (sequences with a known activity profile in selected cell lines). At the design stage, we fix technological and biological constraints (GC 30–40%, restriction motifs/sites, prevention of unwanted patterns) and ensure library diversification to provide models with sufficient training examples.

KEY MILESTONES OF THE PERIOD

- Approved SOPs and library design list; signed contracts with laboratories and services;
- Laboratory SOP validated: acceptance criteria documented, responsible persons appointed;
- Assessment of cycle cost ranges and all possible technological risks;
- Basic release of the technology platform interface and API.

BUSINESS BLOCK

Business application

Creation of basic infrastructure and a pipeline that can be used for research collaborations or grant programmes.

Target audience

Academic laboratories, research institutes, venture bioincubators.

Potential revenue ($)

Undetermined – from grants and research contracts.

Monetisation type

Research Access, grant funding, joint research.

Strategic value

Demonstration of platform readiness, TRL improvement and investment attractiveness.

PHASE 2. FIRST MPRA CYCLES AND MODEL ENSEMBLE

61-135 DAYS

SCIENTIFIC DESCRIPTION

The service laboratory conducts MPRA cycles, unloading, validation, and bioinformatic analysis of the results. At the same time, we **validate the method**: we repeat the mini-run under the same conditions (inter-replicate reproducibility), create **different levels of viral load** and a 'stress matrix' (time/density) to confirm that the laboratory pipeline is **reproducible** and fits within the specified thresholds. Obtaining data on the relationship between sequences and activity in specified cell lines. Submitting this data to the ensemble of models (with parallel selection of the most suitable architectures). Training models and selecting the next sequences for testing (* see note below).

(*) Sequences are selected based on their high activity in pathological cells and lack of activity (or minimal activity, similarly to the presence of suppressive activity) in normal cells.

The best sequences will be thoroughly analysed by models and returned to the design with amendments that the models will make based on the experience of previous experiments. The first experiment (first cycle) will thus be conducted pseudo-randomly, in the absence of the necessary data, but this is a reflection of AI&ML model training, albeit applicable to a biological system – conceptually, models are initialised with random 'weights', the values of which are selected during training cycles – 'epochs'. In our mechanism, the role of 'weights' will be played by DNA sequences, which will change during training, setting the direction of experimental

work towards the shortest path to the best working options, which guarantees a reduction in the number of necessary experimental cycles and, accordingly, the total cost of the work.

The results are automatically uploaded to the platform, forming raw and processed tables, quality control charts, reproducibility charts, model versions, and lists of the next candidates.

We are preparing the following sequence designs for laboratory verification cycles and model retraining.

Technology patenting procedure

At the same time, we are launching a **provisional** application (USA) for the platform methodology **Design → Test (MPRA) → Learn (models) → Redesign** as a related technological process (iterative cycle with data and model feedback). The description includes:

- cycle architecture (design constraints, barcode scheme, acceptance/restart criteria, reproducibility metrics);
- algorithmic principles of selection/redesign (model features/scoring, candidate generation rules, and filters);
- data integration and model versioning in the platform;
- application options (cell lines/oncology contexts, delivery vectors).

The goal is to establish a **priority date** for the "design-test-training-redesign" cycle as a technology, without prematurely disclosing the composition-of-matter for specific sequences. We are obtaining a patent that will allow us to start business partnerships.

KEY MILESTONES OF THE PERIOD

- Quality control of experimental cycles completed;
- First version of the model uploaded to the platform.
- List of candidates for the selected task.
- Correlation of bioreplicas above the target threshold, control behaviour as expected.
- Acceptance protocol published in the datarum: what is considered ‘accepted/restart’.
- **Provisional application submitted**: methodology of the Design→Test→Learn→Redesign platform cycle described; priority date recorded.
- Data trace published on the platform: library version **4000×50 bp**., replication results, QC/reproducibility metrics, list of candidates for redesign.

BUSINESS BLOCK

Business application

Creation of the first AI model for predicting the activity of regulatory sequences; formation of a dataset suitable for commercial licensing.

Target audience

Biotechnology start-ups, companies in the field of functional genomics, synthetic biology laboratories.

Potential revenue ($)

Undetermined – licensing of the first model, analytical access.

Monetisation type

Model licensing, SaaS access to the AI platform.

Strategic value

Creation of the company's first IP asset, proof of technology viability.

PHASE 3. MVP AND BD START

136-180 DAYS

SCIENTIFIC DESCRIPTION

We are transitioning from parallel to individual testing. For top candidates (benchmark: 10–25 sequences), we produce individual mini-constructs with a reporter (luciferase/fluorescent variants according to the protocol) and perform analysis ("clone assay") in a "tumor/normal" pair. The selection of top candidates (10–25) is based on the results of several 4000×50 bp cycles and subsequent model redesign. At this stage, we are interested in two classes of parameters:

1. **effect on/off** – frequency of activation in tumour cells at minimum/zero activity in normal conditions;
2. **effect stability —** reproducibility when varying the **minimum viral load**, incubation time, sowing density, and other technological factors (mini 'stress tests' according to an agreed matrix of conditions).

Clone analysis also serves as **final validation of the laboratory pipeline**: we confirm signal stability when changing viral load/time/density and set thresholds for subsequent cycles.

At the same time, we calculate confidence intervals for bioreplicas, assess variability, and confirm that control sequences behave as expected. We separately record any signs of 'cross-activation' in the norm and enter such constructs into the platform's 'red list' (automatic warnings in the interface during subsequent design).

Based on the validation results, we generate an **MVP report**: we describe the methodologies (constructs, primer sequences, RT-PCR/sequencing protocols

(Real Time PCR, real-time polymerase chain reaction), analysis parameters), provide source tables (raw and processed), quality control charts, on/off graphs, stress test results, statistical conclusions, and recommendations for the next design cycle. The report also includes a **pipeline validation act**: a list of checks, metrics achieved, and readiness status. This report, along with the complete data set, is published **within the platform**: access via a secure data cabinet (datarum), differentiation of rights (view/export/comment), model versions, and automatic connection to relevant projects.

There, we also prepare 'replicators' – ready-made experiment templates that can be re-run by partners (including reagent lists and instrument parameters).

Next, we **launch BD (Business Development)** in the platform logic. We compile a **Pilot One-Pager**: we formulate the pilot task in terms of a platform project (1-2 iterations of **Design → MPRA → Learn**, duration of **8–12 weeks per cycle**, clear composition of output artefacts – validated sequences with on/off metrics, reports, exportable tables, recommendations on vector/cell context).

We add basic economics (price formula per cycle, options – organoids/primary), SLA (Service Level Agreement) in terms of timing and quality, as well as conditions for data and IP (composition-of-matter on sequences for us; for the partner – a licence for winning sequences upon meeting the criteria) to the document.

We prepare a package of materials for initial contacts: a brief presentation of the platform (5-7 slides), a technical summary (quality control, on/off, model accuracy), demo access to the account (with anonymised data), NDA/MSA and SOW (Statement of Work) templates for the pilot study.

KEY MILESTONES OF THE PERIOD

- Confirmed **on/off** for leading sequences at the individual clone level; mini-stress test report (minimum viral load/time/density) with conclusions on stability.
- **Pipeline 'accepted'**: laboratory part validated, metrics fixed as thresholds for the future.
- **MVP package** deployed on the platform: datarum, model versions, export, access control; 'red/yellow lists' of sequences.
- **BD** launched: Pilot One-Pager, set of materials and demo office ready; first technical descriptions for pilot studies assigned.

BUSINESS BLOCK

Business application

Transition from a research product to a demo MVP. Used for negotiations with potential partners and investors.

Target audience

Pharmaceutical companies, venture investors, accelerators, and start-up incubators.

Potential revenue ($)

$50k–$150k for demo access or trial licences.

Monetisation type

Sale of demo access, preparation of pilots, consulting agreements.

Strategic value

Creation of a product showcase, start of commercial activity, formation of first deals.

PHASE 4. PILOT AND QUALITY GROWTH

6-12 MONTHS

SCIENTIFIC DESCRIPTION

We move from the 'pilot' library to the production library: expanding the pool to 25–40 thousand sequences (inserts ~200 bp), adding natural benchmarks – regulators known from literature/public datasets and 'cold' negatives for calibration – and strengthening the design modes (gradient SeqParam, evolutionary, stochastic) with technological constraints (GC, motifs, restriction sites) and mandatory diversification. The transition from initial insertions of 50 bp to production libraries of 25-40 thousand sequences (insertions ~200 bp) is due to the need to capture longer motifs/site combinations and increase model resolution; technical additives and QC standards are retained. Target metrics at this stage: AUC (Area Under Curve, accuracy) of the model 0.85–0.90, for leading sequences – confirmed on/off with a 15–20-fold difference in the 'tumour/normal' pair. Before starting the pilot, we repeat the **control calibration** of the pipeline on the reference panel: we confirm that the laboratory metrics are stable and meet the thresholds.

We launch the first paid pilot as a managed project within the platform. The partner has their own account:

- Views model versions (marks, metric differences), reports, and quality control charts.
- Views the pipeline quality log: a summary of the latest calibrations and control runs;
- Monitors dashboards on/off for leading sequences and stress matrices (MOI/time/density);

- receives implementation recommendations (vector, cell context, expression thresholds) and ready-made exports (CSV/Parquet/PDF, API tokens);
- works according to agreed SOW and SLA (8–12 weeks/cycle, composition of artefacts, acceptance criteria).

The entire data flow (raw/processed tables, shortlist of leading sequences, validation protocols) is published in the data room with differentiated rights and audit logs. Based on the results of the pilot study, there is a standard fork: subscription to the platform + licensing of the best sequences (terms are fixed in advance).

BUSINESS BLOCK

Business application

Launching paid pilot projects with pharmaceutical companies or biotech companies, testing the effectiveness of the technology in real-world conditions.

Target audience

Pharmaceutical corporations, R&D centres, clinical laboratories.

Potential revenue ($)

$450,000 per pilot. Annual potential (6 pilots/year): ≈ $2.7M/year.

Monetisation type

One-off pilot contracts, validation reports, access to analytical tools.

Strategic value

Initial confirmation of market demand and ROI. Creation of cases for scaling.

PHASE 5. SCALING AND LICENSING

12-18 MONTHS

SCIENTIFIC DESCRIPTION

For sequences that have undergone clonal analysis, in agreement with the partner, we raise extended validation: we connect '**organs-on-a-chip**', **organoids** (three-dimensional cultures) and primary cells. This ladder is necessary to demonstrate the transferability of the effect beyond cell lines and to prepare materials for the partner's regulatory and technological agenda. For each new level of validation, we update the **SOP and acceptance checklists** (from lines to volumetric cultures and primary cells). This formalises the **transferability of the laboratory pipeline** beyond classical cell lines.

In parallel, we prepare a **PCT application** for the best sequences with a **composition-of-matter** formula (we prioritise designs with stable on/off and confirmed specificity). In the platform for these sequences, we are opening separate projects with a full set of artefacts: extended validation reports, model versions, summary dashboards, and audit logs.

IP / US non-provisional patenting. We are filing a patent application in the US for the Design→Test (MPRA)→Learn→Redesign platform methodology with provisional priority (phase 2). The application covers the method, system, and machine-readable medium; the description includes process and quality control parameters (libraries ~4000×50 bp in early cycles and ~200 bp in production, barcodes/UMI, MOI/time/density condition matrices, acceptance/restart criteria). We upload the registration number and receipt, formulas, drawings, IDS, current model versions, and SOPs to the report.

Commercial goals for this horizon remain pragmatic: two active pilot programmes in operation and at least one licence for sequences (structure: upfront + milestones ± royalty by agreement).

At the same time, **MRR (Monthly Recurring Revenue) from subscriptions** is growing: partners continue to use the platform to **obtain activities**, **new designs**, and access to **data/reports** in the data room, which establishes us as infrastructure rather than a one-time contractor.

BUSINESS BLOCK

Business application

Launch of IP licensing mechanism and commercialisation of the platform through subscriptions.

Target audience

Pharmaceutical corporations, large biotech companies, gene editing laboratories.

Potential revenue ($)

Licensing: ≈ $4M/year (2 licences × $1.5–2M).

Subscriptions (ARR): ≈ $2.2M/year.

Libraries: ≈ $1.5M/year.

Total commercial potential after 24 months: $8–10M/year.

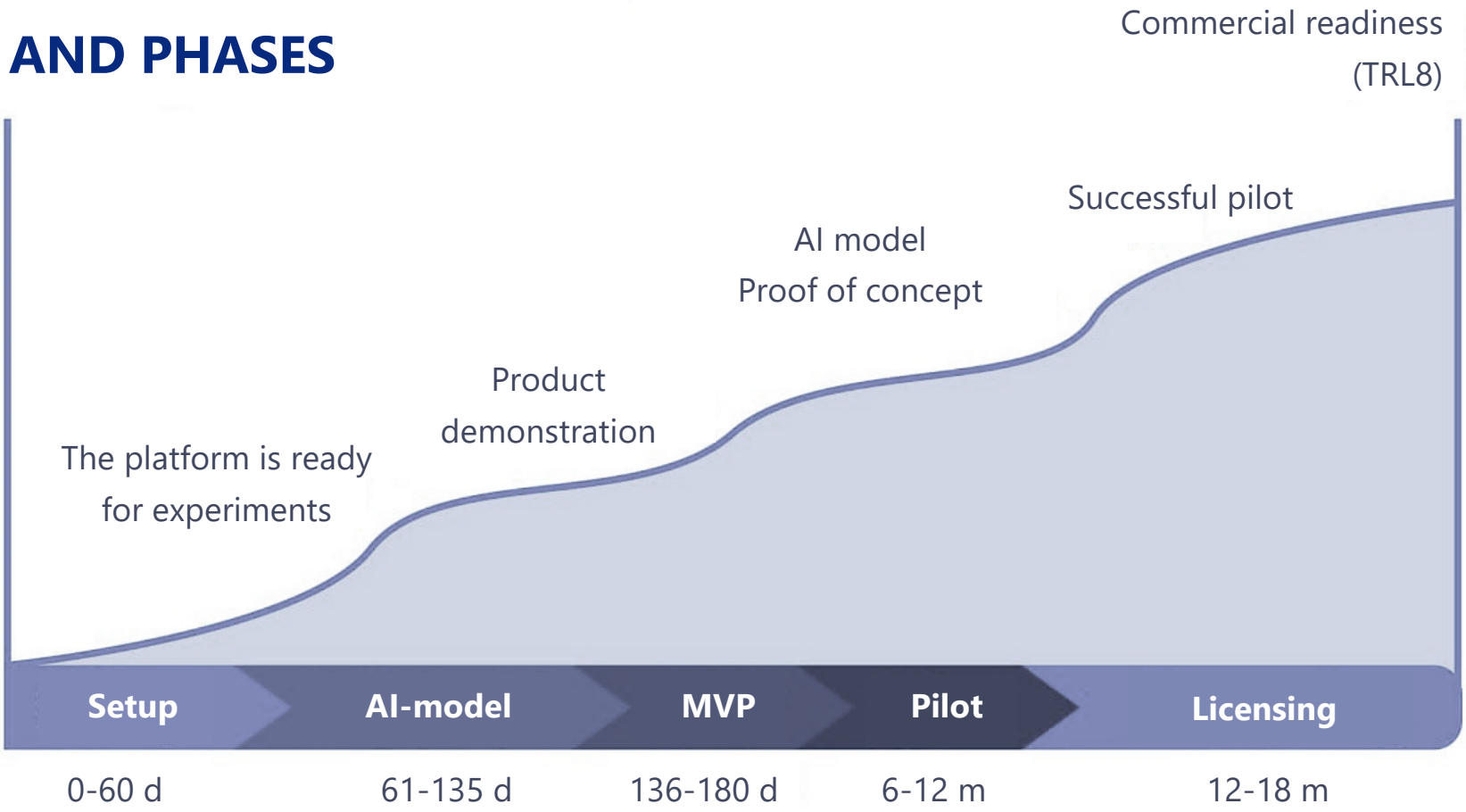
Monetisation type

Upfront payments, licence royalties, ARR subscriptions.

Strategic value

Transition to sustainable ARR, capitalisation growth, market consolidation as a technology platform.

PROJECT STRUCTURE AND PHASES



https://www.oncoswitch.ai

malika@oncoswitch.ai
vitalii@oncoswitch.ai

Table of key milestones for each stage.

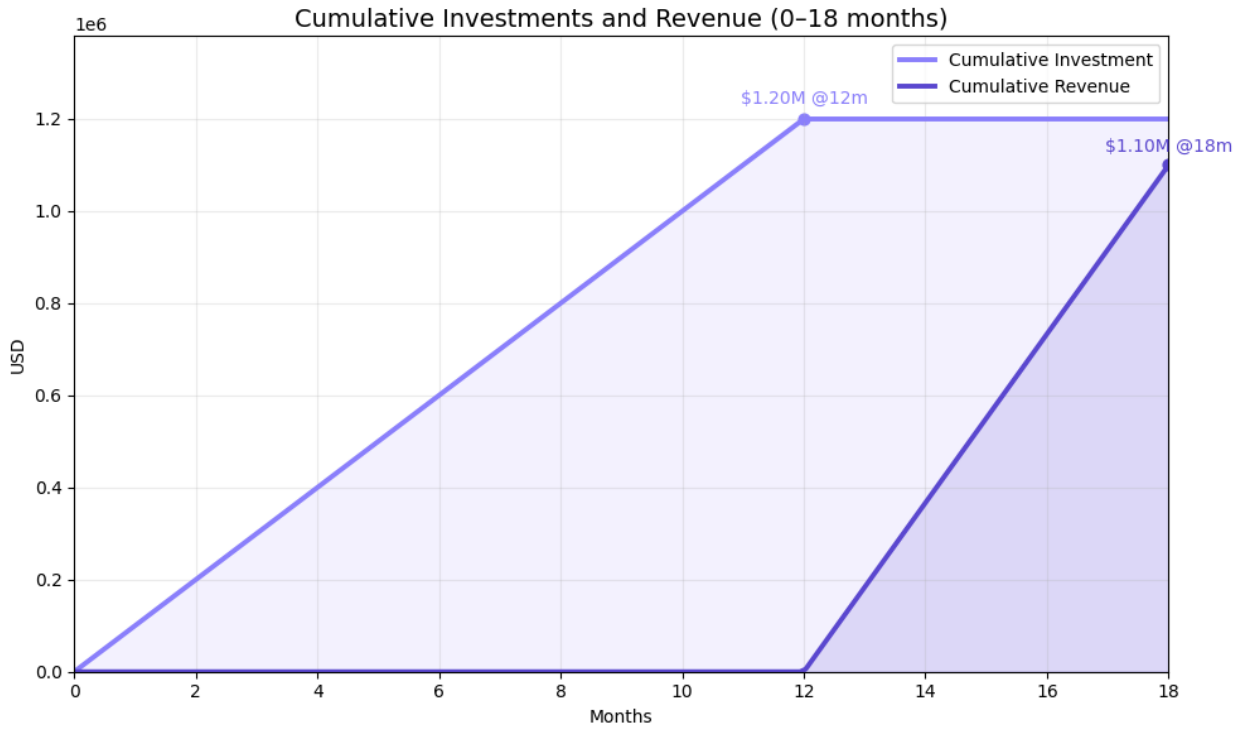
Phase	Science	Engineering	Commerce	Regulatory/operations
Preparation (0–60 days)	Tumor/normal pair; MPRA validation; control panel and QC thresholds.	SOP fixed; basic interface and API; loading/analytics pipelines.	Agreements with service laboratory; cost and risk assessment.	Responsible persons and acceptance criteria have been appointed; restart regulations.
AI model (61–135 days)	First MPRA cycles; reproducibility; sequence→activity profiles.	Ensemble trained; auto-import into platform; shortlist of candidates.	First version of the model in the office; reports and QC cards; next library design.	In the data room — acceptance protocol and QC thresholds; data/model versions.
MVP (136–180 days)	Top candidate clone assay; confirmed "on/off" and stability.	MVP report (methodologies, tables, QC, stress tests); red/yellow lists.	Pilot One-Pager; package of materials (presentation, demo, NDA/MSA/SOW).	The pipeline has been validated and standardized; access and auditing have been configured.
Pilot (6–12 months)	Production library 25–40k; AUC targets 0.85–0.90; on/off 15–20×.	MOI/time/density dashboards; calibration log; exports and API.	First paid pilot; subscription/licensing terms agreed.	SOW/SLA in progress; full audit in the data room; SOP update.
Licensing (12–18 months)	Extended validation: organoids/primary; effect transferability.	Separate leader projects; summary panels; new level checklists.	≥2 pilots; ≥1 license (upfront + milestones ± royalty); subscription MRR growth.	PCT application (composition-of-matter); updated SOP/checklists; package for regulatory agenda.

FINANCIAL MODEL OF THE PROJECT

From costs to income

CYCLES 10 cycles at $50k $500k	**ADDITIONAL EXPENSES** $700k	**ANGEL stage** $1.2M
REVENUE after 24 months $8–10M/year	**ROI** 18 months 5.5×–7×	**ARR POTENTIAL** subscriptions only $2.2M/year

Chart illustrating cumulative costs and revenues over 18 months:



https://www.oncoswitch.ai
malika@oncoswitch.ai
vitalii@oncoswitch.ai

Risk management

The measured “strength” of the insert depends on the technical details of MPRA

Risk. The measured activity of the sequence depends on the vector and minimal promoter, whether it is a plasmid or integration into the genome, and the length and environment of the insert. Changing the context changes the signal, which affects the ranking of sequences and “specificity” .

How we prevent it. To avoid this, we keep the vector and promoter unchanged within the series; we fix the length of the insert; we include “empty”/inert inserts as a baseline background – controls; we validate the leading sequences in the second context (plasmid → integration) and on an alternative cell model; we keep the same metrics at all steps [1, 3, 5].

Library: synthesis and representation

Risk. Uneven oligonucleotide representation, synthesis errors, extreme GC, repeats, and hidden motifs narrow the dynamic range and introduce biases.

How we prevent it. To reduce risks, we check the library before and after sequencing (we set clear QC and SOPs for suppliers); at the design stage, we limit GC windows, homopolymers, repeats, and prohibited sites; we use a mixed pool (pure random + motif-directed design); we duplicate critical pools with a second supplier; we include redundant barcodes and always normalize by RNA/DNA; we make replicates at the lot level [1, 2, 3, 6].

Transferability of leads from cell lines to primary/organoids (MPRA)

Risk. Sequences that are strong in cell lines lose selectivity in more “realistic” models.

How we prevent it. To avoid this, we follow the validation "ladder": line → organoids/primary cells; identical metrics at each stage; stress tests; we record negative results and exclude them from further design [1, 2, 3].

AI retraining and fixation on technical features

Risk. The model uses technical features (vector, length, barcode patterns, batch) instead of regulatory logic.

How we prevent it. We make separate splits by designs, batches, and lines; at the very end, we also include a separate independent test library on the best sequences (we check it both in biology and in silico); we exclude any technical markers from the features; For each model release, we form a mandatory interpretation: importance maps, in silico mutagenesis, motif analysis [1, 4].

Ignoring the "regulatory grammar" model

Risk. Failure to take into account flanks, order, and distances between motifs worsens the transferability of predictions and the quality of design.

How we prevent it. We use syntax-sensitive architectures (convolutions + attention/dilations); libraries where order/distances/flanks are systematically varied; target mini-pools for testing identified rules; a report on flank change robustness [2, 4].

Design optimization with a model leads to undesirable sequence regions.

Risk. Searching for maximum activity results in GC spikes, homopolymers, repeats, critical sites, and reduced specificity.

How we prevent it. We use a multi-criteria goal—we maximize activity in the target cell and simultaneously penalize off-target and leaks; We impose strict limits on GC windows, repeats, and prohibited motifs. After generation, we filter candidate sequences. We compare them with native and synthetic benchmarks. We confirm the results with experiments on a mini-pool [2, 3]. We also calibrate the model and its decision-making—we bring probabilities to real risk (temperature calibration, ensembles), check the quality of selection with PR@k and hit rate metrics on independent mini-pools, and purposefully select data through active learning in areas of high uncertainty [1, 7].

References:

1. La Fleur A, Shi Y, Seelig G. **Decoding biology with massively parallel reporter assays and machine learning**. Genes Dev. 2024 Oct 16;38(17-20):843-865. doi: 10.1101/gad.351800.124. PMID: 39362779; PMCID: PMC11535156.
2. Zong-Heng Fu, Si-Zhe He, Yi Wu, Guang-Rong Zhao, **Design and deep learning of synthetic B-cell-specific promoters**, *Nucleic Acids Research*, Volume 51, Issue 21, 27 November 2023, Pages 11967–11979, https://doi.org/10.1093/nar/gkad930
3. Gosai SJ, Castro RI, Fuentes N, Butts JC, Kales S, Noche RR, Mouri K, Sabeti PC, Reilly SK, Tewhey R. **Machine-guided design of synthetic cell type-specific *cis*-regulatory elements**. bioRxiv [Preprint]. 2023 Aug 9:2023.08.08.552077. doi: 10.1101/2023.08.08.552077. Update in: Nature. 2024 Oct;634(8036):1211-1220. doi: 10.1038/s41586-024-08070-z. PMID: 37609287; PMCID: PMC10441439.
4. de Almeida BP, Reiter F, Pagani M, Stark A. **DeepSTARR predicts enhancer activity from DNA sequence and enables the de novo design of synthetic enhancers**. Nat Genet. 2022 May;54(5):613-624. doi: 10.1038/s41588-022-01048-5. Epub 2022 May 12. PMID: 35551305.
5. Inoue F, Kircher M, Martin B, Cooper GM, Witten DM, McManus MT, Ahituv N, Shendure J. **A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity**. Genome Res. 2017 Jan;27(1):38-52. doi: 10.1101/gr.212092.116. Epub 2016 Nov 9. Erratum in: Genome Res. 2018 May;28(5):766.3. doi: 10.1101/gr.237321.118. PMID: 27831498; PMCID: PMC5204343.
6. Gordon MG, Inoue F, Martin B, Schubach M, Agarwal V, Whalen S, Feng S, Zhao J, Ashuach T, Ziffra R, Kreimer A, Georgakopoulos-Soares I, Yosef N, Ye CJ, Pollard KS, Shendure J, Kircher M, Ahituv N. **lentiMPRA and MPRAflow for high-throughput functional characterization of gene regulatory elements**. Nat Protoc. 2020 Aug;15(8):2387-2412. doi: 10.1038/s41596-020-0333-5. Epub 2020 Jul 8. Erratum in: Nat Protoc. 2021 Jul;16(7):3736. doi: 10.1038/s41596-020-00422-z. PMID: 32641802; PMCID: PMC7550205.
7. Rafi AM, Nogina D, Penzar D, Lee D, Lee D, Kim N, Kim S, Kim D, Shin Y, Kwak IY, Meshcheryakov G, Lando A, Zinkevich A, Kim BC, Lee J, Kang T, Vaishnav ED, Yadollahpour P; Random Promoter DREAM Challenge Consortium; Kim S, Albrecht J, Regev A, Gong W, Kulakovskiy IV, Meyer P, de Boer CG. **A community effort to optimize sequence-based deep learning models of gene regulation**. Nat Biotechnol. 2025 Aug;43(8):1373-1383. doi: 10.1038/s41587-024-02414-w. Epub 2024 Oct 11. PMID: 39394483; PMCID: PMC12339383.

Resource planning

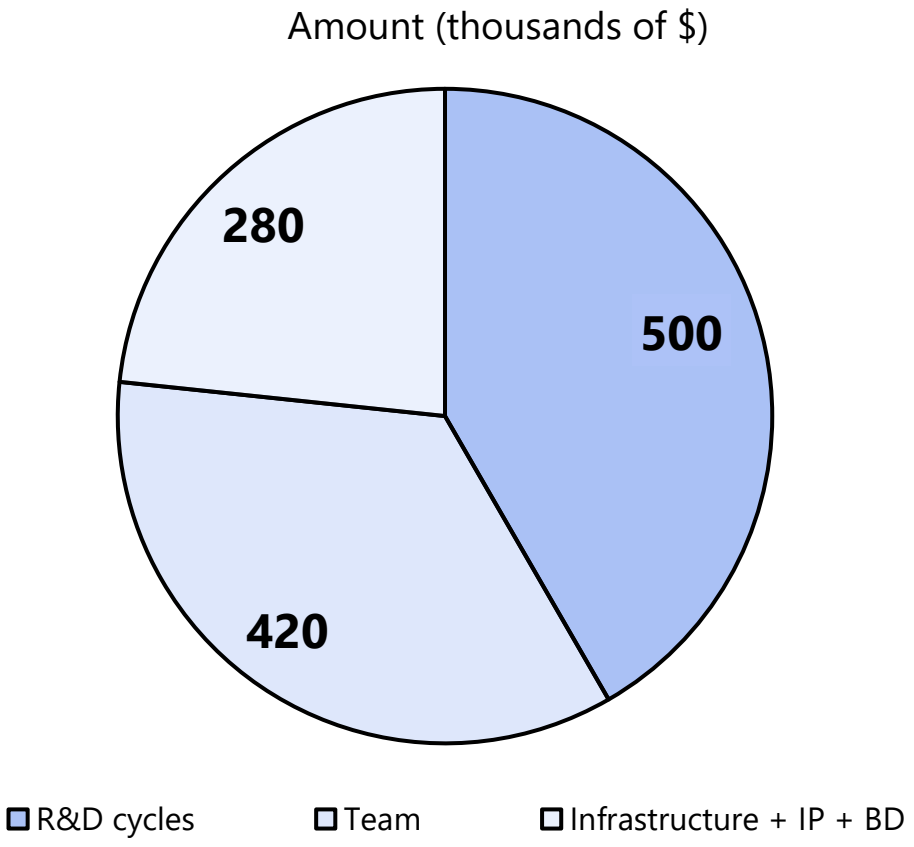
Category	Amount ($)	%
R&D cycles	500k	42%
Team	420k	35%
Infrastructure + IP + BD	280k	23%

We keep the core of our team compact. The CEO is responsible for developing partnerships and deals, while the CTO is responsible for science and technology (AI/genomics, model architecture, and MPRA methodology). At the level of methodology and reliability, we are strengthened by our scientific advisor, **Prof. Dr. Dmitry Mikhaylov** *(Applied AI; 100+ peer-reviewed publications; nine books; NUS & Khalifa University)*: he formalizes data governance and quality standards, designs reliability/stability and safety metrics, leads validation and robustness testing, and translates industrial-grade ML practices into our biomedical stack – turning research into a dependable product. We maintain a focused core team built around essential scientific, operational and business roles. The CEO (1 FTE) and CTO (1 FTE) onboard in month 1 and lead strategy, partnerships, fundraising, and all scientific and technical development — including MPRA architecture, experimental validation, and IP-critical R&D. Wet-lab activities start immediately with the Molecular Biologist (0.75 FTE, month 1), covering MPRA execution, viral transduction, NGS preparation and QC. Day-to-day coordination of labs, vendors, timelines, compliance and documentation is led by the Operations Manager (1 FTE, month 1). In month 4, the Legal/IP Consultant (0.5 FTE) joins to handle provisional filings, FTO analysis, NDAs, contracts and regulatory support. In month 5, the AI/ML Scientist (0.5 FTE) is

added to develop MPRA data models, automate QC and drive sequence-design iteration.

Wet-lab execution beyond the internal molecular biology role remains outsourced under MSA/SOW with a dedicated PI (Principal Investigator) and project manager on the service-lab side, enabling scalable MPRA throughput without expanding permanent headcount.

Also at the start – a minimum of two suppliers (main + backup) to duplicate critical stages: oligonucleotide synthesis, lenti-packaging, sequencing. We make sure that SOPs are agreed upon, quality control thresholds are fixed, and that regular audits are performed (checklist in the datarum).



Budget and program scope (up to 10 cycles; $1.2M)

The economics of each MPRA cycle remain predictable.

Direct costs per cycle are in the **$50,000** range with optimized volumes; in an "extended" configuration, when we grow libraries and add in-depth tests, the bill can rise to **$75,000**.

On top of lab costs are standard operating expenses: team planning, cloud and tools, legal unit and IP, and office. We **spread patent payments across quarters (Q1/Q2/Q4)** to avoid a one-time "peak" in payments and cash flow shortages.

Hiring and contractors.

- **Q1 (Launch Phase)**
  CEO (1 FTE) and CTO (1 FTE) start in month 1; the Molecular Biologist (0.75 FTE) initiates wet-lab operations. MSA with the main service lab is executed, with backup suppliers secured. The Operations Manager (1 FTE) coordinates vendors, timelines, documentation, compliance, and procurement.
- **Q2 (Infrastructure & IP Phase)**
  The Legal/IP Consultant (0.5 FTE) joins in month 4 for provisional filings, FTO, NDAs, and partner contracts. CTO and Operations handle platform scaling and documentation.
- **Q3–Q4 (Scale-Up & Pilots Phase)**
  The AI/ML Scientist (0.5 FTE, starting month 5) develops MPRA models and QC automation. Operations expands into reporting, data-room management, and pilot support as activity increases.

Investor Value / Business Model

Product

We sell the **platform as a service**: a user-friendly interface and API, projects with a full set of reports, models with version history, and a library of data-driven switches.

Each MPRA cycle **adds value** – new data emerges, metrics improve, and the set of leading sequences grows – while simultaneously **reducing the cost** of the next cycle thanks to a better model and streamlined pipeline.

Revenue Model:

Source	Average bill	Frequency	Potential ($)
Pilots	$450k	6/year	2.7M
Subscriptions (SaaS)	$220k	10/year	2.2M
Licenses	$1.5–2M	2/year	4M
Libraries	$300k	5/year	1.5M

Total: $8-10 million/year after 24 months.

Monetization

1) Pilot projects within the platform

Standard managed cycle: **$450k per pilot, 8–12 weeks** depending on the partner's context, with optional **organoids/primary** models for extended validation. The partner receives full visibility into data, QC, on/off metrics, model versions, and implementation recommendations.

Example unit economics of a pilot project:

- Basic cycle COGS: **$150–180k** (MPRA, calibrations, replicates).
- Price: **$450k** per pilot.
- Gross margin: **~60–65%** (excluding extended validations).
- Organoids/primary options: **+$25–80k** to COGS – added only when required for licensability of top sequences.

The margin is important, but the strategic value is in **the data generated and the licensing-grade** leads emerging from each pilot.

2) Platform Subscription (SaaS)

$220k/year for access to the interface, API, updated models, versioned datasets, and libraries; usage-based extensions (data volume, additional projects, expanded dataroom) are available. This generates MRR/ARR and stable recurring revenue that scales with the breadth of the product and the depth of integration.

Upsell logic rem: Pilot → Core Platform → modules (indication-based libraries, advanced checks, private models, additional seats/roles).

Within the dashboard this is natural: the partner selects the needed components and immediately gains access to data/tools.

3) Sequence Licenses (IP Monetization)

When a sequence demonstrates validated selectivity and stability, the commercial **IP framework is activated: upfront $1.5–2M, milestones $5–10M, royalties 2–4%**. This provides **strong upside** and high profitability.

The structure remains flexible: indication, territory, exclusivity, development timelines, and rights reversion in case of inactivity.

4) Co-development with pharma

Full-scale co-development programs with shared risks and shared data flow: the partner brings platforms, vectors and disease models; we provide the design-MPRA-model cycle.

The outcome is early validation in difficult contexts (AAV, oncoviruses, CAR-T, TCR) and accelerated licensing of lead sequences generated within such programs.

Additional Product Lines

Product expansion

12–24 months

- ❑ New libraries (liver, lung, gliomas)
- ❑ Co-development programs
- ❑ Partner databases
- ❑ API licenses

We also sell switch libraries: ready-made, validated modules tailored to specific indications. A full library package is priced at **$300k**, providing partners with immediately deployable regulatory elements for internal R&D. Libraries generate direct revenue and expand our installed base – increasing the number of downstream licensing opportunities.

In addition to libraries, we license developed models and datasets: aggregated MPRA data, validated feature sets, and “frozen” model weights/configurations packaged as platform artifacts. This is a separate, premium component for R&D teams that need a reliable starting point for their own pipelines or internal model development.

Conclusion

We combine the power of artificial intelligence and living biology in a single, honest cycle: AI designs, the lab tests, and together they find the most precise DNA switches. The vision we're working toward is simple and bold: a treatment that activates only where needed, remaining silent in healthy tissue. Without unnecessary pain. Without paralyzing fear. With a normal human emotion – hope.

We're not trying a one-time trick – we're building a platform that can change the rules of cancer therapy. A platform where each cycle adds data, precision, and confidence. We need partners who believe in science, value emotion, and choose life.

Intelligence may be artificial, but human life will always remain real. Let's launch the world's first platform for generating DNA switches together and take a step toward therapy you can truly trust.

Investment Impact Summary

Indicator	Before Investment	After 18 Months
Cycles	0	10
Validated Switches	0	5
Clients	0	3
Partners	0	2
ARR	0	$2.2M+
Patents	0	≥2 PCT

Appendix A.

Glossary of Terms

DNA switch – a regulatory DNA sequence that turns on the expression of a therapeutic gene in tumor cells and is "silent" in normal tissue; a target artifact of the OncoSwitch platform.

"50 bp + technical additives" – the length refers to the active regulatory insert; additionally, the construct includes service elements (primer sites/linkers, barcodes/UMI, adapters) that are not considered part of the active sequence.

The volume of "~4000 sequences per cycle" is the size of unique active inserts; the number of reads/cells is calculated specifically for QC purposes and statistical power.

MPRA / lenti-MPRA – massively parallel reporter assay; tests thousands of DNA sequences simultaneously, often via lentiviral delivery (lenti). Output: "sequence → activity" in a "tumor/normal" pair.

Pilot – a paid, managed project (8-12 weeks/cycle) within the platform: design → MPRA → analysis → report → recommendations. Provides case/data for a subscription or licensing.

License – transfer of rights to use selected sequences (composition-of-matter), typically structured as follows: upfront + milestones ± royalty.

Dataroom – a secure data room: reports, QC cards, model versions, audit logs, access control.

Organoids / primary – 3D tissue cultures / primary cells; stages of extended validation of signal transferability beyond cell lines.

AUC – area under the receiver operating characteristic curve; model target values 0.85–0.90 during the pilot phase.

TRL – Technology Readiness Level; target commercial readiness state ~TRL-8.

SOP – standard operating procedure; specifies QC thresholds, acceptance criteria, and restart criteria.

SLA / SOW – service level agreement / specification of pilot work (deadlines, artifacts, metrics, acceptance).

PCT – international patent application; for leading sequences (composition-of-matter).

MOI – multiplicity of infection (virus per cell) in lenti experiments; one of the "stress matrix" parameters.

QC – quality control (library quality, barcode saturation, bioreplica correlation).

ROI / ARR – return on investment / annual recurring revenue. In the model: ROI ≈ 1.4× for 18 months; ARR potential ≈ $1.5 M/year.

Appendix B.

Financial settlements by phases

Phase	Duration	Basic expenses (benchmarks)	Income / Benefits	Key results
Preparation	0-60 days	Team, platform launch, MVP library: **$140k**	—	SOPs, laboratory contracts, basic interface, risk assessment.
MPRA + AI Models	61-135 days	2 MPRA cycles (~$50,000/cycle), analysis, cloud: **$260k**	—	First "sequence → activity" profiles, 1st version of the model, shortlist of candidates.
MVP & BD Launch	136-180 days	Clone essays, MVP report, BD package: **$120k**	Demo/Test Access: **$50–$150k**	On/off verification, dataroom, pilot package (One-Pager, NDA/MSA/SOW).
Pilot	6-12 months	Production library, calibrations, 3 MPRA cycles: **$340k**	1 Paid Pilot: **$450k**	AUC 0.85–0.90; confirmed leaders 15–20× "on/off".
Scale & licenses	12-18 months	IP, 4 MPRA cycles, validation – **$350k**	Licenses + SaaS + libraries: **$8–10M/year** trajectory	≥2 pilots in operation, ≥1 license; PCT for leaders; MRR/ARR growth.

18-month outlook: total costs ≈ $1.2M (up to 10 MPRA cycles × ~$50k), initial revenue ≈ $0.8–1.2M from early pilots and demo access, with post-18-month annualised revenue potential of ≈ $8–10M/year (pilots, subscriptions, licenses, libraries).

Budget structure: R&D cycles ≈ $500k; team ≈ $420k; infrastructure + IP + BD ≈ $280k.

Formulas:

ROI = (Revenue − Costs) / Costs; ARR = MRR × 12

Appendix C.

Financial calculations by cycle

C1. Cost structure (COGS, basic configuration)

Phase	What's included	Amount
P1. Library Design & Synthesis	AI design + oligonucleotide pool	$21,000
P2. Vector Cloning & Library Construction	Cloning into lentivirus vector, pDNA baseline	$3,750
P3. Lentiviral Packaging & Transduction	Packaging, concentration/titer, infection	$4,000
P4. Cell Treatment & Analysis	Incubation window before NK isolation	$800
P5. Reporter Assay & Sequencing	RNA/DNA → cDNA/UMI → library prep → sequencing	$16,700
P6. Data Analysis & Reporting	Calculations, QC, report	$2,000
PM/logistics/buffer	Delivery, express services, minor contingencies	$1,750
	TOTAL	**$50,000**

C2. Laboratory steps and timeline (8–12 weeks/cycle)

Design → synthesis/assembly → lentiproduction → transduction and incubation → extraction → NGS → analysis and QC → report and recommendations for the next design.

C3. Cycle KPIs

- Barcode saturation and proportion of "callable" constructs (SOP threshold).
- Bioreplicate correlation (RNA/DNA normalization).
- Positive/negative pass control.
- Model AUC improvement; number of "on/off" leaders ≥ N×.
- Cycle time (weeks), proportion of stages without restarts (according to the acceptance protocol).

C4. Economics and ROI per cycle (for the pilot)

For clarity, an MPRA cycle is a **cost component inside a pilot**, not the pilot price itself.

Baseline cost of one MPRA cycle: ≈ **$50,000** (18-month budget actual). At a $50k cost and a $50k internal transfer price, **gross margin = 0%, ROI = 0 (breakeven).**

Cycle ROI formula:

ROI = (Price − COGS) / COGS

At $50k / $50k → ROI = 0

With **COGS = $50k**:

- Margin **40%** → price **$83k.**
- Margin **50%** → price **$100k.**
- Margin **55%** → price **$111k**

COGS (with options)	Break-even price	Price for 40% margin	Price for 50% margin
$65k	$65k	$108k	$130k
$80k	$80k	$133k	$160k
$110k	$110k	$183k	$220k