Rivercell

Dossier candidature
Pépinière Paris Santé Cochin

Juillet 2026

Nom de la société : Rivercell (Blossom Life Sciences)
Nom du dirigeant : Yann Fleureau
Dossier présenté dans le cadre de la pépinière

CONFIDENTIEL

Fiche de synthèse

Nom de la société	RIVERCELL (Blossom Life Sciences)
Adresse du siège social	23 rue du Départ, boîte 37, 75014 Paris
Date de création	24/10/2023
Forme juridique	SAS
Représentant (nom et fonction)	Yann Fleureau, CEO
Secteur d'activité	Biotechnologies
Produit / service	Plateforme de génération de données de perturbation cellulaire
Caractère innovant	Augmentation du débit de génération de données d'environ 400 fois par rapport aux technologies existantes, avec un appariement image + transcriptome à l'échelle de la cellule unique et la capacité d'explorer des combinaisons de perturbations (chimiques × génétiques).
Marché	Marché de la recherche en drug discovery (préclinique) ; marché de l'AI Virtual Cell (modèles d'IA en biologie cellulaire)
Composition du capital	Fondateur : 100% ND, 87,5% FD Employés : 9% FD Advisors : 3,5% FD (hors conversion des BSA Air)
Effectif actuel	6
Effectif à 4 ans	50
CA 2024 (dernier CA connu)	0 € (société en phase de R&D, pré-commerciale)
Levées de fonds réalisées	Pre-seed : 460 000 € (février 2025) Finalisation d'un tour de seed de 20m€ - trois termsheets reçues, 4m€ déjà reçus sur le compte en banque.
Financements publics obtenus	970 000 € d'aides : BFTE (90 k€ perçus), FPI (30 k€ perçus), WILCO (60 k€ perçus), iLab (370 k€ dont 259 k€ perçus), Innov'up Leader PIA (420 k€).
Sources de revenus	n/a (pré-commercial)
Trésorerie au 01/07/2026	4 040 954.58€
Surfaces demandées	Bureaux : 140 m² - Laboratoire L1 : 120 m² - Laboratoire L2 : 38 m²
Collaborations envisagées avec les équipes Cochin	Collaboration avec les biotechs installées (accélérer leur recherche grâce à la plateforme) ; collaboration avec les équipes de recherche de l'hôpital (tests de cellules primaires sur la plateforme).

I. Présentation de l'entreprise

Coordonnées

Société	Rivercell (Blossom Life Sciences)
Nom	Evain
Prénom	Hélène
Fonction occupée	COO
Adresse	25 rue de Tolbiac, 75013 Paris
Téléphone	06 24 24 06 20
Mail	helene@rivercell.ai

Secteur d'activité
Biotechnologies et IA appliquée à la biologie.

Équipe
Voici l'organigramme de la société à date, ainsi qu'une présentation des membres de l'équipe.


Yann Fleureau
Co-founder & CEO


Eric Durand
Co-founder & CSO


Hélène Evain
COO


Chérif Adouama
Material Science Eng.


Peta Bradbury
Lead Biologist


Aurélie Deleforge
Head of Partnerships

Yann Fleureau (Co-founder & CEO, Rivercell)

- Diplômé de l'École Polytechnique (X2010) et de UC Berkeley (2013).
- Co-fondateur et ancien CEO de Cardiologs Technologies, une scale-up créée en 2014 et acquise par Philips en 2021. Cardiologs est le leader du marché de l'IA appliquée au diagnostic de maladies cardiaques, permettant le diagnostic de plus de 2 millions de patients par an.
- Investisseur et board member : il a investi de l'ordre de 2m€ dans les domaines de la medtech (Moon Surgical, Reliev, Sunrise, Sonio…) et de la biotech (Bon Vivant Food, Surge…).
- Co-fondateur et président du projet associatif Numina, qui développe des modèles d'intelligence artificielle dans le domaine du raisonnement et des mathématiques.

Eric Durand (Co-founder & CSO, Rivercell)

- Doctorat en statistiques appliquées et génétique des populations de Grenoble INP - UGA (2009), lauréat du prix AAAS Newcomb Cleveland.

- Chercheur postdoctoral à UC Berkeley (2009-2011) : développement de méthodes statistiques d'inférence de l'histoire évolutive humaine.
- Computational Biologist chez 23andMe (2011-2015) : développement de techniques de machine learning sur des données génomiques et conception des fonctionnalités d'ascendance de la plateforme.
- Director Data Science en oncologie chez Novartis (2015-2022) : développement et application de méthodes d'IA et de bio-informatique sur données multi-omiques, imagerie et pathologie digitale, sur l'ensemble du pipeline de développement de médicaments.
- SVP Data Science puis Chief Data Science Officer chez Owkin (2022-2026), aujourd'hui Strategic Advisor.
- Co-fondateur de Bioptimus (depuis 2024, non opérationnel), conseil en IA appliquée à la découverte de médicaments (Fractional CSO, AI advisor, board member).

Hélène Evain (COO, Rivercell)

- Diplômée d'HEC (2017).
- Expérience de 3 ans en conseil en stratégie puis de 4 ans chez Cardiologs en tant que *Business Operations Manager* (préparation de la Série B et travail sur la vente à Philips) puis *Business Development Manager* (responsable du lancement d'un nouveau produit)
- Co-fondatrice et trésorière du projet associatif Numina, qui développe des modèles d'intelligence artificielle dans le domaine du raisonnement et des mathématiques.

Aurélie Deleforge (Head of Partnerships, Rivercell)

- Diplômée de l'École Polytechnique (X2012) avec un master en Biologie et de l'Université de Cambridge (2016) avec un master en Bioscience Enterprise.
- Expérience de 4 ans chez Bionest Partners en tant que consultante et manager auprès d'entreprises biopharmaceutiques.
- Directrice de l'unité *Remote Patient Monitoring* chez Cardiologs (2020-2022) : développement de l'offre, définition de la vision produit et commercialisation aux États-Unis et en Europe.
- Directrice des Partenariats chez Owkin (2022-2025) : responsable de la stratégie commerciale et R&D pour la Pharma, et gestion opérationnelle des projets de fourniture de données.

Peta Bradbury (Lead Biologist, Rivercell)

- Doctorat en biologie cellulaire et moléculaire de l'Université de Sydney (2016)
- Plus de 10 ans d'expérience en recherche, spécialisée en biologie du cancer, maladies respiratoires et criblage de molécules (test de composés inhalés, modèles in vitro et précliniques in vivo).
- Scientifique senior et cheffe de projet au Woolcock Institute of Medical Research à Sydney (2017-2020) : pilotage de 2 essais cliniques (biomarqueurs du cancer du poumon et sevrage tabagique), obtention des approbations réglementaires et éthiques.
- Scientifique senior à l'Institut Curie (2020-2024) : direction d'un projet préclinique in vivo en oncologie puis développement d'un projet de biologie du développement intégrant mécanique des jonctions cellulaires, bio-ingénierie et CRISPR/Cas9.

Chérif Adouama (Material Science Engineer, Rivercell)

- Ingénieur et chercheur en chimie organique.
- Master en synthèse organique et chimie des molécules bioactives et Doctorat en synthèse organique à l'Université Claude Bernard à Lyon.
- Deux post-doctorats au CEA (radiomarquage en microfluidique) et à la Fondation des Sciences et du Patrimoine (développement de dosimètres à rayons X.
- Expériences en tant qu'ingénieur chercheur chez Alvéole et Alchimedics.

Rivercell a également réuni un collège d'experts regroupant des personnes ayant une expérience solide dans le domaine de la biologie et de la microfluidique digitale (*electrowetting*).

Cyril Delattre nous apporte son expertise en microfluidique digitale développée au CEA et chez Advanced Liquid Logic (Illumina).

Cédric Viry nous apporte son expertise en science des matériaux appliquée à la microfluidique digitale développée chez Volta Labs.

Valérie Taly (CNRS) nous apporte son expertise en microfluidique et un large réseau et une bonne connaissance de l'écosystème de la recherche académique en France et à l'étranger.

David Bikard (Institut Pasteur) nous accompagne dans l'identification d'applications pour la technologie.

Troy Lionberger nous accompagne dans l'identification d'applications pour la technologie, ainsi que dans l'évaluation des différents modèles économiques grâce à son expérience chez Berkeley Lights.

Robert Sebra (Mount Sinaï) nous accompagne dans l'identification d'applications pour la technologie.

Thibaut Mercey nous apporte son expérience dans le développement d'entreprise et sa connaissance de l'écosystème d'acteurs de la biotech en France.

En plus de son collège d'experts, Rivercell a accès à un réseau de potentiels partenaires et conseils, tels que DNA Script, Twist Bioscience, Ginkgo, Eligo Bioscience, Numi, Orakl Oncology, Surge, Neoplant, MethysDx.

Rivercell a réalisé en février 2025 une levée de fonds sous forme de BSA Air de 460 000€, afin de s'entourer d'investisseurs ayant une expérience pertinente pour le projet (biologie, biotech, électronique, etc.). Une liste non-exhaustive des personnes qui soutiennent financièrement le projet est présentée ici.



Structure juridique, date de création & décomposition actuelle du capital
La société est une SAS, immatriculée le 24 octobre 2023.

L'unique actionnaire de Rivercell est Yann Fleureau.
Il détient 85% de la société via sa holding Wakaru Capital, dont il est l'unique actionnaire, et 15% en direct.
La société a mis en place un plan de BSPCE pour ses salariés et de BSA pour ses advisors :

- Les salariés actuels disposent de 9% (*fully diluted*) de la société
- Les advisors disposent de 3,5% (*fully diluted*) de la société

La société a également réalisé une levée en BSA Air en février 2025 de 460 000€. Tous les *business angels* ont été rassemblés dans un SPV.
La société est en train de finaliser un tour de seed de 20m€, et a reçu trois termsheets.

GLOSSAIRE

Terme	Définition
AI DD (Artificial Intelligence for Drug Discovery)	Utilisation de l'intelligence artificielle pour accélérer la découverte de nouveaux médicaments, en analysant des données biologiques et chimiques pour identifier des cibles, optimiser des composés et prédire l'efficacité des candidats-médicaments.
AIVC	Modèle de machine learning visant à représenter et prédire le comportement d'une cellule (définition complète page 9) Dans le document, nous parlons de cellule virtuelle, AIVC, AI Virtual Cell
Cell painting	Technique d'imagerie cellulaire à haut débit utilisant des colorants fluorescents pour visualiser différentes structures cellulaires, permettant l'analyse phénotypique des effets de divers composés sur les cellules.
CRO (Contract Research Organization)	Entreprise spécialisée dans la prestation de services de recherche notamment pour l'industrie pharmaceutique et biotechnologique, couvrant des domaines comme les essais précliniques, les essais cliniques et l'analyse des données.
Drug Discovery	Dénomination de l'étape initiale du processus de découverte, de développement et de mise sur le marché de nouveaux médicaments, comme défini dans la Figure 2.
Machine learning	Branche de l'intelligence artificielle qui permet aux algorithmes d'apprendre à partir de données et d'effectuer des prédictions ou des classifications.
NGS	Technologie de séquençage de l'ADN et de l'ARN à haut débit permettant l'analyse rapide et précise du matériel génétique. Utilisée en recherche biomédicale pour identifier des mutations, analyser l'expression génique et accélérer la découverte de cibles thérapeutiques.
Readout	Mesure expérimentale qui peut être obtenue grâce à de la lecture optique ou à des moyens analytiques avancés tels que le séquençage ou la spectrométrie de masse. **Bulk** : Mesure réalisée sur un ensemble de cellules simultanément, dont le résultat correspond à une moyenne de la population. Plus simple à produire que le single cell, mais masque les différences entre cellules individuelles. **Single cell** : Mesure réalisée au niveau de chaque cellule individuelle plutôt que sur une population moyennée, permettant de capturer l'hétérogénéité réelle entre cellules et le comportement propre à chacune. **Monomodal** : Se dit d'un readout ne capturant qu'un seul type de mesure à la fois (par exemple uniquement le transcriptome, ou uniquement la morphologie). **Multimodal** : Se dit d'un readout combinant plusieurs types de mesure capturés simultanément sur une même cellule (par exemple empreinte moléculaire, morphologie et comportement fonctionnel), révélant des informations invisibles à chaque modalité prise isolément.
Wet lab	Laboratoire où sont réalisées des expériences biologiques et chimiques nécessitant l'utilisation de liquides, de réactifs et de cellules, par opposition aux laboratoires "secs" dédiés à l'analyse informatique et aux simulations.
Workcells	Systèmes automatisés intégrant plusieurs modules robotiques pour exécuter des protocoles expérimentaux de biologie de manière automatique

II. Description du projet

Origine du projet

Rivercell est née de deux constats et de deux convictions.

Conviction 1 : L'intelligence artificielle va révolutionner notre compréhension de la biologie.

L'IA est une opportunité technologique et scientifique majeure de notre époque. La performance des réseaux de neurones s'améliore avec la quantité de données d'entraînement et les premières applications en biologie ont déjà donné lieu à des résultats remarquables :

- AlphaFold et Rosetta, des modèles révolutionnaires de prédiction de structures protéiques, ont été récompensés par le prix Nobel de chimie[1] en 2024, et de nouveaux modèles de fondation émergent chaque mois à partir de données publiques et privées.
- L'Institut ARC[2] a publié Evo[3], un modèle de fondation d'ADN entraîné sur 80 000 génomes archéens et bactériens, représentant 300 milliards de nucléotides-tokens. Les premiers résultats sont déjà prometteurs (découverte de nouvelles variantes d'enzymes Cas9 par exemple) bien que la base de données d'entrainement reste limitée par rapport à la diversité génomique existant dans la nature, sans parler de la diversité transcriptomique, et en comparaison avec les LLMs (GPT-4 entrainé sur environ 13 trillions de tokens). L'utilisation des ressources de calcul reste également modeste face aux modèles de langage : Evo a nécessité 10^{22} FLOPS (environ $100,000 par pré-entraînement) contre 10^{25} pour GPT-4.

Nous sommes convaincus que nous n'en sommes qu'aux prémices de l'IA générative appliquée à la biologie, et en particulier, que la prochaine grande opportunité pour l'IA est la biologie cellulaire.

La publication de Bunne et al.[4] en décembre 2024 a cristallisé la vision de l'AI Virtual Cell (AIVC) : un système neuronal multi-échelle et multimodal, capable de prédire les réponses cellulaires à toute perturbation et dans tout contexte biologique.

Constat 1 : Les datasets utilisés pour l'entraînement des modèles d'AIVC actuels sont monomodaux, mono-perturbations et mono-*readout* : ils manquent de richesse. Cela tient aux infrastructures de génération de données en *wet lab*, non conçues pour produire des données à l'échelle des besoins de l'IA.

Les performances des modèles d'AIVC actuels reflètent directement le manque de richesse des données disponibles.

Les larges datasets existants (Tahoe-100M[5], X-Atlas/Pisces[6]) sont monomodaux, monoperturbations et dominés par des cancer cell lines, alors que l'IA exige des données combinatoriales, multimodales et longitudinales. Comme le résume une chercheuse dans le domaine de l'AIVC, ces données « *do not suffice to produce models that beat linear baselines […] and explore a very small portion of the perturbation space* ». Essentiellement issues de jeux publics, les données d'entraînement du domaine sont structurellement déficientes sur quatre axes :

1. La **modalité** (quasi exclusivement RNA ; readouts phénotypiques, protéiques, spatiaux et métabolomiques absents) ;
2. Le **type de perturbation** (single-agent uniquement ; perturbations combinatoires inexplorées) ;
3. La **temporalité** (snapshots statiques plutôt que trajectoires longitudinales) et

[1] https://www.nobelprize.org/prizes/chemistry/2024/press-release/
[2] https://arcinstitute.org/
[3] Nguyen & al., Sequence modeling and design from molecular to genome scale with Evo, Science Vol 386 Issue 6723
[4] Bunne C et al. How to build the virtual cell with artificial intelligence. Cell. 2024;187(25):7045–7063. doi:10.1016/j.cell.2024.11.015
[5] Zhang, Ubas, de Borja, Svensson et al. Tahoe-100M: A large-scale single-cell perturbation atlas. bioRxiv. 2025.
[6] Wang B et al. (Xaira Therapeutics). X-Cell: Scaling Causal Perturbation Prediction Across Diverse Cellular Contexts via Diffusion Language Models. bioRxiv. 2026. doi:10.64898/2026.03.18.712807

4. Le **contexte biologique** (surreprésentation des cancer cell lines, couverture minimale des cellules primaires et tissus).

Ces lacunes constituent la raison centrale de l'échec des modèles d'AIVC à généraliser, comme le confirment les benchmarks récents[7]. Les infrastructures de génération de données de référence reposent aujourd'hui sur des solutions d'automatisation robotisées dont le coût atteint plusieurs millions de dollars pour une chaîne entièrement automatisée de bout en bout - peu flexibles et structurellement incapables de fournir le volume et la richesse de données nécessaires à l'entraînement de modèles d'AIVC performants.

Constat 2 : L'innovation à l'interface de l'IA et de la biologie se concentre sur le développement de modèles et non la génération de données, qui est pourtant un maillon essentiel et stratégique dans la chaîne de valeur.

De nombreux acteurs ont émergé ces dernières années dans le domaine de l'intelligence artificielle appliquée à la biologie. Certains investissent dans la génération de données, à l'image de Recursion, qui a intégré le *cell painting* dans ses processus, ou d'A-Alpha Bio, dont la plateforme permet une production à grande échelle de données sur l'interaction des protéines. Cependant, aucune de ces entreprises n'a introduit une innovation de rupture fondamentalement capable d'augmenter exponentiellement le débit de production de données, qui est pourtant la matière première indispensable au développement et à la validation de modèles. Les entreprises développant des modèles d'IA appliquée à la biologie que nous avons rencontrées se procurent leurs données soit en contractualisant avec des CRO classiques, soit en assemblant leur propre *wet lab* à partir de technologies sur étagère.

Conviction 2 : La microfluidique digitale est une technologie arrivant à maturité dont le potentiel reste entièrement à exploiter pour révolutionner la génération de données en biologie. Il y a un intérêt stratégique à développer cette capacité en France.

Aujourd'hui, le développement de puces de microfluidique digitale appliqué à la réalisation d'expériences de biologie est un domaine principalement actif aux Etats-Unis (Berkeley Lights / Beacon racheté par Bruker en 2023 après avoir levé $400m, Volta Labs $20m Series B) et au Royaume-Uni, et plus précisément à l'université de Cambridge dont ont émergé Nuclera ($87m Series C en 2024/26), Acxel ($10m Series A) et Lightcast ($50m Series B en 2024 et $27m en 2026).

L'état de l'art existant permet de réaliser plusieurs milliers d'expériences de biologie moléculaire et cellulaire en parallèle sur des puces allant jusqu'à 1m de pixels. Tous les acteurs existants se positionnent en tant que vendeurs d'instruments et font le choix de spécialiser leur équipement sur un protocole spécifique, ce qui les empêche d'exploiter rapidement le plein potentiel de la technologie, à la fois en matière de débit et de flexibilité de programmation.

Notre approche se différencie ainsi :

- Développement d'une puce conçue pour passer à l'échelle la génération de données de perturbations cellulaires, capable de traiter en parallèle plus de 350 000 cellules et 2 000 perturbations (et combinaisons de perturbations). L'importance de la donnée de perturbation cellulaire pour l'entrainement de modèles d'IA et en particulier, de cellules virtuelles (AIVC) est expliquée dans la section suivante ;
- Intégration end-to-end de la puce avec les instruments de mesure (*readouts*);
- Exploitation de la technologie en interne plutôt que sa commercialisation comme outil de laboratoire, afin de tirer pleinement parti de la flexibilité et de la programmabilité de la puce.

Nous sommes convaincus de l'importance stratégique de développer une technologie de rupture en France dans le domaine de la génération de données en biologie. Cet atout sera un facteur clé pour assurer, sur le long terme, la compétitivité française dans le domaine de l'intelligence artificielle appliquée à la biologie.

[7] DenAdel A et al. Evaluating the role of pretraining dataset size and diversity on single-cell foundation model performance. Nat Methods. 2026. doi:10.1038/s41592-026-03120-y

Description du produit, service ou procédé, stade de développement à ce jour

Rivercell développe une infrastructure de génération de données en biologie cellulaire, différenciée et scalable.

Notre plateforme permettra d'offrir simultanément :

- des **perturbations combinatoires à grande échelle** (nous visons un débit 400 fois supérieur aux technologies existantes) ;
- des **readouts multimodaux**, couplant l'imagerie en continu à la transcriptomique single-cell pairé à l'échelle de la cellule ;
- des **assays « arrayed »,** qui associent sans ambiguïté chaque perturbation à la cellule mesurée, contrairement aux approches pooled où ce lien doit être reconstruit a posteriori ; et
- des **données longitudinales**, grâce à un design expérimental itératif et résolu dans le temps, compatible avec l'active learning.

Nous sommes convaincus qu'il s'agit d'une couche manquante essentielle pour rendre possibles les modèles de cellule virtuelle.

Infrastructure de génération de données (wet lab)

Rivercell développe un hardware propriétaire qui permet une échelle sans précédent. Notre infrastructure repose sur une puce combinant microfluidique digitale à grande échelle et computer vision. Elle permet de traiter 2 000 conditions par run (200 cellules par condition), en appliquant aussi bien des perturbations externes (petites et grandes molécules) qu'internes (génétiques), ainsi que leurs combinaisons, sur cellules adhérentes et en suspension. Les cellules sont cultivées et exposées aux perturbations en « bulk », avec environ 70 cellules par gouttelette ; les cellules adhérentes sont maintenues dans leur condition physiologique d'adhérence grâce à des *microcarriers*. Le *readout*, lui, demeure *single-cell* et pairé à l'échelle de la cellule : chaque cellule est suivie de manière individuelle et multimodale — imagerie en time-lapse (40X, fond clair et 4 canaux fluorescents) couplée à la transcriptomique (scRNA-seq), pour capturer l'hétérogénéité des réponses à travers plusieurs niveaux d'organisation de la cellule.

Autour de cette technologie, Rivercell repense l'ensemble du workflow de préparation cellulaire au sein d'unités de production que l'on appelle BPU (Bio Processing Unit). Un BPU réunit deux plateformes de préparation Rivercell, un incubateur conçu pour héberger les puces et capturer l'imagerie en direct et un séquenceur Ultima Genomics. Cette intégration end-to-end, de la perturbation jusqu'au readout, permet à un seul BPU de produire chaque année 250 millions de cellules multimodales, soit 1,2 million de perturbations uniques.

Entrainement de modèles de machine learning dans le domaine de l'AI Virtual Cell

En parallèle du développement hardware, Rivercell travaille sur des modèles de cellules virtuelles.

En biologie cellulaire, un modèle de cellule virtuelle est un modèle d'IA qui apprend à représenter le fonctionnement d'une cellule à partir de données biologiques massives, afin de prédire sa réponse à une perturbation donnée (gène inhibé, molécule, dose). Il vise à simuler le comportement cellulaire *in silico*, et donc à remplacer une partie des expériences de paillasse par des prédictions computationnelles. À la différence du texte qui a alimenté les LLM, la donnée pour entrainer des modèles de cellule virtuelle n'existe pas à l'échelle requise : elle doit être produite expérimentalement. L'espace des perturbations (gènes × composés × doses × combinaisons × types cellulaires × temps) étant combinatoirement immense, la capacité à générer ces données à haut débit et haute résolution devient l'étape limitante - et le véritable créateur de valeur - pour entraîner des modèles capables de généraliser vers des perturbations jamais observées.

Notre travail sur le développement de modèles de cellules virtuelle bénéficie donc directement du wet lab scalable, flexible et à latence quasi nulle que nous construisons (et qui intègre notre hardware propriétaire). Grâce à cet avantage, nous créons un environnement de reward et une feedback loop très efficace pour entraîner des modèles de *machine learning*.

En couplant les modèles et le wet lab, nous avons, à terme, pour ambition de créer une usine à actifs thérapeutiques et d'accélérer drastiquement l'identification et la validation de candidats médicaments.

Stade de développement à ce jour

- Réalisation d'une preuve de concept de la technologie propriétaire, permettant de réaliser des expériences de biologie cellulaire ;
- Mise en place de la chaîne d'approvisionnement pour les dépôts de couches minces et les sources de lumière ;
- Trois brevets en cours de dépôt ;
- Construction (en cours) d'un Scientific Advisory Board structuré autour de trois missions (plus de 20 profils pertinents ont déjà été identifiés, contractualisation en cours avec Charlotte Bunne) :
    1. Construire la crédibilité scientifique autour du projet, et dérisquer la roadmap hardware ;
    2. Construire des datasets d'évaluation et définir les métriques et benchmarks contre lesquels évaluer les modèles, de sorte que ces évaluations soient validées par l'industrie et représentatives des enjeux de l'early drug discovery ;
    3. Établir des relations et des partenariats avec l'industrie pharmaceutique.
- Constitution (en cours) d'un premier dataset de données multimodales (généré à partir de technologies existantes), pour initier le développement de modèles d'AIVC et les premières études sur les *scaling laws*.

Caractère innovant

La solution développée par Rivercell rompt avec les approches actuelles de génération de données biologiques en faisant passer le screening « arrayed » à une toute autre échelle : non pas davantage de cellules par condition, mais un nombre de conditions étudiées individuellement sans précédent. Couplée à une imagerie en continu directement sur la puce, elle ouvre la voie à des données multimodales à l'échelle single cell.

Elle repose sur la microfluidique digitale (DMF), une technologie émergée dans la recherche à la fin des années 1990 qui permet d'adresser et de manipuler individuellement des gouttelettes de liquide par l'application de champs électriques : chaque goutte se comporte alors comme un micro-puit piloté numériquement. Longtemps freinée par des problèmes de fiabilité et par des brevets verrouillant - notamment ceux d'Illumina, dont les titres maîtres expirent en 2026, la DMF a fortement progressé grâce aux avancées en sciences des matériaux et en électronique, et offre aujourd'hui un potentiel majeur de réduction des coûts, des consommables et des déchets, ainsi qu'une flexibilité de programmation et un débit accru.

Si plusieurs sociétés ont déjà industrialisé la DMF pour la préparation d'échantillons ou les essais cellulaires, aucune n'en exploite encore tout le potentiel, tant en débit qu'en flexibilité.

Le projet Rivercell ambitionne d'aller bien au-delà de l'état de l'art actuel en relevant trois défis majeurs :

- Exploiter pleinement l'échelle permise par la microfluidique digitale : l'augmentation de la taille des puces et du nombre de gouttelettes manipulables soulève des défis techniques majeurs, en particulier la génération d'une force suffisante pour réaliser des opérations de *liquid handling* complexes, comme le *split* (séparation d'une goutte en deux).
- Atteindre la résolution cellulaire grâce à des algorithmes de *machine learning* et de *computer vision* qui permettent, avec une imagerie en continu directement sur la puce, de maintenir l'identité cellulaire tout au long du protocole.

- Structurer nativement les données produites afin de les rendre directement exploitables pour l'entraînement de modèles d'IA, et permettre la mise en place d'une *feedback loop*.

En combinant DMF à très grande échelle et computer vision, Rivercell construit une capacité de génération de données fortement différenciée par rapport au marché.

Au-delà de la plateforme hardware, Rivercell porte une ambition tout aussi innovante sur le plan du *machine learning* : entraîner des modèles de cellules virtuelles appelés à inverser le paradigme actuel de la biologie, en la faisant passer d'une science aujourd'hui à 90 % expérimentale et 10 % computationnelle, à l'équilibre inverse.

Propriété industrielle

Rivercell travaille actuellement sur le dépôt de trois brevets pour protéger les innovations sur la plateforme propriétaire de microfluidique digitale et sur le protocole expérimental pour la génération de données de perturbations cellulaires avec, pour chaque cellule, le lien entre les données d'imagerie et les données de scRNA-seq.

Les innovations développées par Rivercell sont propriétaires.

La propriété intellectuelle de Rivercell sera structurée autour d'un portefeuille de brevets pour les procédés brevetables et d'un système de secrets industriels pour les innovations non brevetables.

La marque Rivercell est en cours de dépôt en Europe et aux États-Unis, ainsi que le nom de domaine associé. Toute discussion avec d'éventuels partenaires est encadrée par des accords de non-divulgation (NDA) et de propriété intellectuelle.

III. Plan de développement

Le plan de développement de Rivercell s'articule autour d'objectifs à court terme et de deux roadmaps en parallèle - hardware et AIVC - ayant vocation à se nourrir mutuellement et à converger.

A court terme nos objectifs sont les suivants :

- Construire des partenariats avec la pharma pour construire des métriques et des datasets d'évaluations sur des use case industriels (Target : 10 datasets) ;
- Construire une équipe top tier en AI/cell bio et reproduire le SOTA des instituts américains (Arc / CZI) et évaluer les modèles sur les use case industriels ;
- Construire un laboratoire qui génère ~50M de single cell multimodales par an à partir de hardware existant sur le marché (nécessite l'achat de plusieurs machines) ;
- Construire notre hardware propriétaire qui permettra de produire 250 millions de single cell multimodales par an par machine ;
- Démontrer la supériorité des données single cell multimodales pour créer l'équivalent d'AlphaFold en biologie cellulaire.

Sur le plan hardware, la roadmap prévoit :

- **La livraison d'un MVP à un an**, atteignant l'échelle cible avec un pairing image / scRNA-seq à l'échelle de la perturbation (et non de la cellule), validé sur lignées cellulaires adhérentes et compatible avec les perturbations à partir de petites molécules et biomédicaments ;
- **La livraison de la plateforme complète à deux ans**, offrant une résolution à l'échelle de la cellule et compatible avec les lignées cellulaires comme avec les *primary cells* et les iPSC, en suspension et en adhérence.

En parallèle, Rivercell conduit un programme de développement de modèles de cellules virtuelles (AIVC) structuré autour de trois objectifs :

- Constituer des métriques et datasets d'évaluation de référence, validés par l'industrie et en particulier la pharma ;
- Établir les scaling laws des modèles d'AIVC, en démontrant notamment l'apport de la donnée multimodale pour justifier l'entrée dans une phase d'hyperscale ;
- Entraîner nos propres modèles d'AIVC, en s'appuyant sur le wet lab automatisé comme source de données et feedback loop.

IV. Étude du marché

Taille actuelle du marché et tendance

1. Fourniture de données pour l'AI drug discovery (AI DD)

Rivercell se positionne d'abord sur le marché émergent de la fourniture de données pour l'AI DD, caractérisé par un fort déséquilibre entre l'offre et la demande de données. En dix ans, ce secteur a levé près de 20bn$ et conduit une vingtaine de candidats-médicaments jusqu'aux phases cliniques. L'essor de l'IA générative a récemment accéléré les investissements et fait émerger de nouveaux acteurs développant des modèles de fondation du vivant (représentations in silico, prédiction de structures et de fonctions protéiques) afin d'accélérer et de fiabiliser chaque étape de la drug discovery.

En témoigne la course aux *billion-cell atlases* : la CZI y consacre 500 M$ sur cinq ans (dont 400 M$ pour la génération de données) via sa *Virtual Biology Initiative*, quand Illumina, le consortium CZI–10x–Ultima ou Parse–Tahoe visent déjà des atlas de l'ordre du milliard à cinq milliards de cellules. Ces atlas restent toutefois largement mono-perturbations et mono-modaux.

Nos échanges avec plusieurs de ces acteurs confirment leur intérêt pour une solution accélérant la génération de données d'entraînement, en particulier les données de perturbations cellulaires. Le plus grand atlas Perturbseq actuel (readout scRNAseq), Tahoe-100M, ne couvre qu'environ 50 000 perturbations uniques, alors que les clients potentiels souhaitent en explorer de l'ordre du million. Constituer 100 000 points de données de toxicité cellulaire avec *readout* optique demande aujourd'hui ~6 mois : Rivercell ramènerait ce délai à quelques jours.

Nous estimons la taille de cette première opportunité à près d'un milliard de dollars, sur la base des levées des 5 dernières années et d'une part de budget allouée aux données de 30 % pour les acteurs purement in silico et de 10 % pour ceux disposant de laboratoires automatisés. Encore à ses prémices, ce marché devrait croître fortement. La capacité ultra-haut débit de la puce de Rivercell ainsi que la richesse des données générées la positionnent idéalement pour l'adresser.

2. Génération de données in vitro pour l'industrie biopharmaceutique

L'infrastructure de Rivercell adresse également le marché des services en *drug discovery* et en pré-clinique, traditionnellement couvert par les CRO. Les deux premiers segments de la chaîne de valeur — Early Discovery et Pré-clinique — représentent 23bn$ en 2023, en croissance d'environ 10 % par an, portés par la tendance à l'externalisation de la R&D. Trois dynamiques renforcent l'opportunité :

- **Essais fonctionnels plus précoces :** les équipes cherchent à trier et à éliminer plus tôt les hits/leads sans effet sur la cible. Le coût élevé des essais (~50 000 cellules par test) limite aujourd'hui leur passage à l'échelle ; un format single-cell en gouttelettes le rendrait abordable.
- **Recul de l'in vivo :** pour des raisons éthiques et de transposabilité limitée à l'homme (ex. immunothérapies TCR), l'usage des modèles in vivo diminue au profit de modèles in vitro comme les organoïdes.

- **Besoin de customisation :** malgré des offres standardisées par type de traitement (petites molécules, anticorps…), chaque client requiert des protocoles adaptés. La flexibilité-par-design de l'infrastructure de Rivercell permet d'y répondre rapidement et à faible coût.

3. Développement d'actifs thérapeutiques

A terme, Rivercell a pour ambition de créer une usine à actif thérapeutique. En capitalisant sur la richesse de ses datasets single-cell et sur la puissance prédictive de ses modèles de cellule virtuelle, la plateforme permettra d'identifier et de valider de nouvelles cibles, puis de concevoir des candidats-médicaments optimisés *in silico*. Rivercell pourra dès lors développer ses propres programmes thérapeutiques ou les valoriser via des partenariats et accords de licence avec l'industrie pharmaceutique.

Définition des marchés cibles 2026/27

Avec sa solution de génération de données de perturbation cellulaire ultra haut débit et haute résolution, Rivercell cible deux marchés à court / moyen terme, qui seront adressés de manière séquentielle :

- Nous envisageons un premier *go-to-market* sur un marché de niche à forte croissance **: la fourniture de données pour les entreprises spécialisées en *AI drug discovery***. Ce marché est actuellement évalué à plusieurs centaines de millions de dollars.
- Nous ciblerons ensuite un marché plus vaste : celui de la **génération de données *in vitro* pour l'industrie pharmaceutique**, dont la valeur est estimée à plusieurs dizaines de milliards de dollars. Ce marché bénéficie d'une dynamique favorable, notamment portée par le développement de nouveaux modèles *in vitro* (organoïdes) et la recherche d'alternatives aux essais *in vivo* chez l'animal.

Étude de la concurrence

Sur le marché de la vente de données de perturbations, le paysage concurrentiel se structure en deux cercles :

- Les **concurrents directs** sont les CRO et plateformes qui basculent vers un modèle data-as-a-service — Ginkgo Bioworks, Revvity, VectorBuilder, Cellecta. Leur limite reste toutefois le débit ou la profondeur — la plupart s'appuient sur des readouts bulk ou mono-modaux.
- Les **concurrents indirects** sont les fournisseurs d'outils et d'instruments (10x Genomics, Cellanome, Element Biosciences, Illumina…) ainsi que les plateformes intégrées (Tahoe, Xaira) qui permettent la génération de données de perturbation cellulaire, qui équipent le marché mais imposent un arbitrage structurel entre échelle et richesse de la donnée : aucun ne combine aujourd'hui haut débit de perturbations *et* lecture multimodale single-cell.

Positionnement du ou des produits, services et avantages concurrentiels

Rivercell développe une infrastructure intégrée — une plateforme hardware de microfluidique digitale couplée à des modèles de cellule virtuelle — destinée à produire, à très haut débit, des données de perturbation cellulaire multimodales et nativement structurées pour l'entraînement de modèles d'IA. Le produit s'adresse d'abord au marché émergent de la fourniture de données pour l'*AI drug discovery*, caractérisé par un fort déséquilibre entre l'offre et la demande, puis au marché des services pré-cliniques traditionnellement couvert par les CRO. A terme, Rivercell a vocation à créer une usine d'actifs thérapeutiques propriétaires.

Le positionnement de la plateforme Rivercell repose sur trois avantages concurrentiels distinctifs :

- Scalabilité, à la fois dans le nombre de perturbations réalisables et dans l'exploration de leurs combinaisons

- Donnée multimodale et longitudinale à l'échelle de la cellule unique, quand les grands atlas concurrents (Tahoe, Xaira) demeurent largement mono-perturbation et mono-modaux.
- Intégration *en closed loop* du *wet lab* et des programmes de cellule virtuelle, un actif dont la valeur se capitalise à chaque cycle et constitue une barrière à l'entrée difficilement réplicable.

V. Étude commerciale ou modèle économique

Modèle économique envisagé

L'infrastructure que Rivercell développe ouvre une trajectoire de création de valeur à plusieurs niveaux, du revenu à court terme jusqu'à la découverte de médicaments.

- **Vente de données expérimentales**. À court terme, l'infrastructure permet de générer des revenus en commercialisant des datasets perturbationnels auprès des acteurs de l'IA et de l'industrie pharmaceutique.

- **Vente de prédictions in silico.** À mesure que les modèles propriétaires de Rivercell gagnent en maturité, la plateforme pourra non seulement vendre la donnée expérimentale, mais aussi générer des prédictions in silico et commercialiser directement ces résultats. Ce glissement de la donnée vers la prédiction ouvre un revenu à plus forte marge.

- **Vente / licence d'actifs thérapeutiques**. Développement de nos propres programmes scientifiques, valorisés par licences et partenariats. Le wet lab propriétaire de Rivercell accélère la phase d'early discovery : coûts réduits, dérisquage plus rapide des actifs, et durée de vie des brevets maximisée. Les brevets moléculaires étant déposés au milieu de l'early discovery, raccourcir cette phase prolonge d'autant l'exclusivité commerciale en aval.

Ces trois modèles économiques ne s'excluent pas : ils coexistent et se renforcent mutuellement. La vente de données à des tiers — laboratoires pharmaceutiques, biotechs, acteurs de l'IA — joue un rôle particulier dans cette dynamique. Chaque jeu de données validé et exploité par un client externe vient étayer la crédibilité scientifique de la plateforme.

Nous envisageons une stratégie modulable, dans laquelle le degré de verticalisation le long de la chaîne de valeur — génération de données, entraînement des modèles, identification de candidats-médicaments, développement jusqu'à l'IND — varie selon les indications. Sur certaines d'entre elles, Rivercell pourra choisir de s'intégrer verticalement et de créer des spin-offs afin de capturer l'intégralité de la valeur générée par la propriété intellectuelle. À l'inverse, sur d'autres indications, Rivercell pourra prioriser la vente de données (générées expérimentalement ou grâce aux modèles d'IA), permettant à des partenaires pharma ou biotech d'accélérer leur progression vers la phase IND.

Actions commerciales, type de clients

Dans le cadre du modèle économique de vente de données, l'objectif de Rivercell est de construire des laboratoires automatisés dès 2027, en ciblant en priorité deux zones géographiques stratégiques : les États-Unis et l'Europe, qui représentent les deux plus grands marchés mondiaux en matière de R&D pharmaceutique et biotechnologique.

Dans la mesure où l'infrastructure de Rivercell nécessite des investissements et un espace limités (3-4m€ et <200m^2 respectivement), la vision à long terme est d'implanter ces infrastructures dans chaque hub local d'innovation, afin de garantir des délais de génération de données rapides et compétitifs. À terme, Rivercell pourrait même envisager de développer des laboratoires sur site pour ses clients majeurs, en offrant un service dédié à l'exploitation de ces installations.

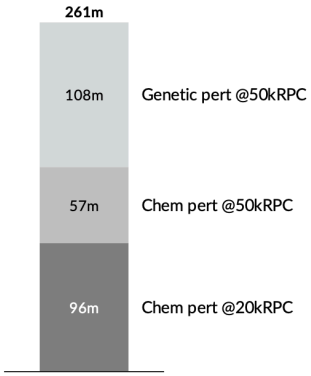
En ce qui concerne le **go-to-market sur ce modèle économique**, nous l'envisageons en trois étapes clés :

1. **Signer les premiers contrats de génération de données (2027)**, activité menée par Aurélie Deleforge (Head of Partnerships).
2. **Construire la crédibilité scientifique de Rivercell (dès cette année) via le Scientific Advisory Board:** Eric Durand travaillera avec nos partenaires et des KOLs pour construire la visibilité scientifique de Rivercell.
3. **Déployer le modèle commercial (à partir de 2028)**

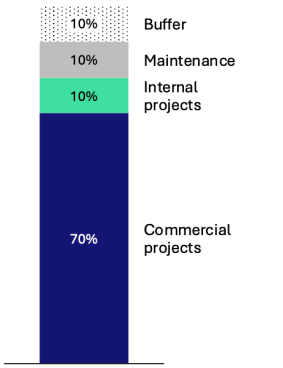
Objectifs commerciaux

L'objectif est de commercialiser chaque Bio Processing Unit à 70% (le reste de la capacité sera utilisé pour des projets internes) et de générer donc 25m€ de revenus par BPU.
Nous avons pour ambition de démarrer la commercialisation des données d'un BPU en 2028.

Lab capacity – 1 BPU I
Number of cells*



Lab utilization rate I
% of total capacity BPU



Proforma P&L I

m€	Proforma
Revenue	**25,4**
COGS	(9,1)
Gross Margin	**16,3**
GM%	*64%*
Lab operations & expenses	(0,9)
Hosting costs	(0,5)
Internal data generation	(1,4)
Sales & Marketing	(5,1)
G&A	(1,8)
OPEX	**(9,7)**
EBITDA	**6,6**
as % of revenue	*26%*

Tableau de financement sur 3 ans

Ci-dessous le plan de financement de Rivercell sur la période 2026-2028 :

Financement	Type	Date	Montant
Seed	Equity	Septembre 2026	20m€ (3 termsheets reçues), 4m€ déjà collectés
Innov'up Leader PIA (contrat signé, 1ère tranche débloquée)	Subvention / Avance remboursable	Q3-26	420k€ (1ère tranche débloquée le 03/07/26)
iLab - 2ème tranche (contrat signé, 1ère tranche reçue)	Subvention	2027	2ème tranche : 111k€ (première tranche déjà reçue, total 370k€)

Autres financements publics	Subvention / Avance remboursable	Objectif : 8m€ sur deux ans. Ci-dessous les programmes identifiés auxquels nous prévoyons de candidater : • ADD : 1.5m€ (Q1-27) • Eurostars : 3m€ (Q1-27) • iDEMO : 1.5m€ (Q2-27) • Innov'up Jeune Pousse Industrielle : 1m€ (Q4-27) • Prêt Nouvelle Usine : 1m€ (Q1-28) PAI (montant à définir)
Série A	Equity	S2-28 (montant à définir, de l'ordre de 50m€)

Partenariats envisagés

Rivercell a déjà tissé des liens avec l'écosystème de la recherche en France et à l'international. Le développement de sa plateforme nécessitera d'approfondir ces relations et d'en construire de nouvelles, structurées autour de trois axes: l'accès à des équipements et plateformes techniques de pointe, la validation biologique sur modèles cellulaires (notamment issus de patients), et l'ancrage précoce auprès de l'industrie pharmaceutique.

a. Accès aux plateformes techniques et équipements de pointe

Le développement de la puce Rivercell repose sur l'accès à des procédés de micro-fabrication avancés, en particulier pour le dépôt de couches minces. Rivercell a engagé à ce titre des contrats de services avec plusieurs plateformes et instituts de référence : le LPICM (École polytechnique) et le Fraunhofer IST, avec lequel Rivercell monte actuellement un consortium dans le cadre d'une candidature pour un financement Eurostars. Au-delà de la micro-fabrication, l'accès à des instruments de pointe en biologie — microscopes haute résolution, séquenceurs — constitue un levier déterminant pour le développement et la caractérisation de la plateforme. La plateforme technique de Cochin (en particulier la plateforme Genom'ic) présente à cet égard un intérêt majeur pour Rivercell, en donnant accès à un parc instrumental de haut niveau au sein d'un environnement de recherche intégré.

b. Validation biologique et construction des jeux de données sur modèles cellulaires

Un enjeu central des deux prochaines années est la validation de la plateforme sur des cellules dérivées de patients (*patient-derived cells*) et l'entraînement de modèles d'IA prédictifs sur ces modèles biologiques cliniquement pertinents. Dans ce cadre, le partenariat avec des services hospitaliers et des équipes de recherche est important, et l'appartenance à l'écosystème Paris Biotech Santé, au contact direct des services cliniques et des équipes de recherche de l'hôpital Cochin, constitue un atout stratégique pour accéder à ces modèles.

c. Ancrage précoce auprès de l'industrie pharmaceutique

Rivercell entend nouer très tôt des liens avec l'industrie pharmaceutique, à travers la constitution d'un *scientific advisory board* réunissant des experts de la R&D pharmaceutique et de la drug discovery. L'objectif est double : valider dès l'amont les benchmarks et les métriques de performance attendus par les futurs clients, et confirmer l'intérêt et la valeur des données générées au regard de leurs besoins réels. Cet ancrage précoce permettra d'orienter les priorités de développement, de sécuriser les premières discussions commerciales et de constituer un vivier de partenariats, construction conjointe de jeux de données, accords de licence, à mesure que la plateforme gagnera en maturité.

VI. Besoins pour la pépinière

CAHIER DES CHARGES :

Nombre de m2 souhaités par typologie de locaux	
Bureaux	**140m2**
Laboratoires L1 avec Sorbonne/sans Sorbonne	**120m2**
Laboratoires L2	**38m2**
Utilisez-vous de l'azote liquide ?	**Non**
Nombre de tanks azote nécessaires	**n/a**
Nombre de frigos, - 80	**1 (+ 1 -150°C)**
Besoins en laverie	**Blouses et évier pour laver l'équipement de laboratoire**
Nombre de postes de travail nécessaires en 2026/2027/2028/2029	**2026 : 25** **2027 : 35** **2028 : 40** **2029 : 50**
Autres besoins spécifiques	**n/a**

BESOIN D'ACCOMPAGNEMENT du DEVELOPPEMENT DE VOTRE ENTREPRISE :

Pour quel type d'actions ?

- Accès aux plateformes techniques de Cochin (notamment Genom'ic)
- Accès aux modèles patients via les équipes cliniques et de recherche
- Appui à la structuration de l'entreprise (recrutement, financements publics, mise en réseau).

Que cherchez-vous en venant vous installer à la pépinière ?

- La proximité clinique et l'accès aux *patient-derived cells*, essentiels à la validation biologique de notre plateforme sur les deux prochaines années
- L'accès à un environnement de recherche intégré et à des équipements de pointe
- L'appartenance à l'écosystème Paris Biotech Santé, au contact de biotechs, de cliniciens et de partenaires potentiels.

Comment envisageriez-vous de participer à la vie du réseau des entreprises qui sont ou s'installeront dans la pépinière ?

En mettant notre plateforme de génération de données au service des biotechs installées pour accélérer leur recherche, en collaborant avec les équipes de l'hôpital autour du test de cellules primaires, et en partageant l'expérience de notre équipe fondatrice (parcours entrepreneurial de Cardiologs à Philips, double expertise IA et biologie).

Accepteriez-vous une mission de mentoring ? Oui

ELEMENTS FINANCIERS ET PIECES A JOINDRE AU DOSSIER

Voir annexes (dossier compressé)

Coordonnées de la personne signataire : Yann Fleureau yann@rivercell.ai 06 87 24 89 10