

# Factorisation Non-Négative de Matrice pour séparation de sources en Imagerie par Bioluminescence préclinique

Erwan DEREURE<sup>1,2</sup>, Elsa ANGELINI<sup>2</sup>, Anikitos GAROFALAKIS<sup>3</sup>, Johanne SEGUIN<sup>4,5</sup>, Nathalie MIGNET<sup>4</sup>, Jean-Christophe OLIVO-MARIN<sup>1</sup>

<sup>1</sup>Institut Pasteur, Université Paris Cité, CNRS UMR 3691, BioImage Analysis Unit, F-75015 Paris, France.

<sup>2</sup>Department Image-Data-Signal, LTCI Telecom-Paris, Institut Polytechnique de Paris, France

<sup>3</sup>Biospace Lab, 11 Rue de Chenival, 95960 Nesles-la-Vallée, France

<sup>4</sup>Université Paris Cité, UTCBS, CNRS, INSERM, 75006 Paris, France.

<sup>5</sup>Université Paris Cité, Plateformes mutualisées de l'institut du médicament LIOPA, CNRS, INSERM, 75006 Paris, France

erwan.dereure@pasteur.fr, elsa.angelini@telecom-paris.fr,  
AGarofalakis@biospacelab.com, johanne.seguin@u-paris.fr, nathalie.mignet@u-paris.fr,  
jcolivo@pasteur.fr

**Résumé** – L'imagerie par bioluminescence est une technologie d'imagerie optique basée sur la détection de photons produits par une réaction biologique entre un substrat et une enzyme dans les tissus vivants. Cette technique d'imagerie, peu bruitée, est utilisée dans de nombreux domaines de la recherche biomédicale et notamment en oncologie préclinique. Le flux de photons émis par une tumeur est lié à son volume et son état métabolique, nécessitant une localisation précise de celle-ci. Certaines expériences impliquent la visualisation de signaux multiples (par exemple 2 tumeurs) sur le même animal ce qui rend leur quantification difficile car la délimitation manuelle de la région d'intérêt (ROI) doit être plus fine et est le plus souvent arbitraire. Dans ce travail, nous étudions l'utilisation de la décomposition factorielle par Factorisation Non-Négative de Matrice (Non-Negative Matrix Factorization, NMF) comme méthode de séparation des sources afin de permettre une quantification plus précise de chaque tumeur sans a priori spatial.

**Abstract** – Bioluminescence imaging is an optical imaging technology based on the detection of photons produced by a biological reaction between a substrate and an enzyme in living tissue. This low noise imaging technique is used in many fields of biomedical research and particularly in preclinical oncology. The photon flux from the tumor is related to the volume and the metabolic state of the tumor. Moreover, some experiments require the visualization of multiple signals (e.g. 2 tumors) on the same animal which makes their quantification difficult because the manual delimitation of the region of interest (ROI) must be more precise and is most often arbitrary. In this work, we investigate the use of Non-Negative Matrix Factorization (NMF) as a source separation method in this context in order to allow a more accurate quantification of each tumor without spatial a priori.

## 1 Introduction

L'imagerie par bioluminescence (BLI) est une technologie d'imagerie optique basée sur la détection des photons produits par l'oxydation d'un substrat (luciférine) injecté par voie systémique [7] (intraveineuse IV - intrapéritonéale IP) ou locale qui réagit avec une enzyme spécifique (luciférase) exprimée dans l'hôte biologique (exemple cellules tumorales modifiées) en présence d'oxygène et d'adénosine triphosphate (ATP). Cette modalité d'imagerie permet le suivi de phénomènes biologiques en oncologie préclinique [5].

Plusieurs paramètres de la BLI peuvent influencer le signal enregistré tels que la profondeur de la tumeur dans l'organisme [12], la sensibilité du système de détection (généralement une caméra CCD), le positionnement de l'animal, la température du tissu ou encore l'anesthésique utilisé [6]. Comme les tissus des mammifères ont peu de bioluminescence naturelle, la BLI semble être une méthode d'imagerie privilégiée pour étudier les phénomènes *in vivo* [13]; en particulier lorsque les tumeurs sont situées profondément dans le corps de l'animal et ne peuvent

donc pas être mesurées de façon non invasive ou sans sacrifier l'animal.

En recherche préclinique de traitements contre le cancer, il est courant d'utiliser le volume de la tumeur comme indicateur de l'efficacité du traitement [11] et c'est cet indicateur que la BLI tente de quantifier de manière non invasive. L'hypothèse de base est que le flux de photons au pic de la réaction est corrélé au nombre de cellules cancéreuses [11, 7].

Pour obtenir cette information quantitative à partir d'une acquisition d'images, la méthodologie généralement utilisée est de dessiner manuellement une Région d'Intérêt (ROI) autour de la zone tumorale dans le logiciel d'analyse d'images intégré au système d'acquisition de bioluminescence [8]. Selon Zilberman et al. [16], comme la bioluminescence n'est pas un signal très bruité, il est plus imprécis de dessiner une ROI trop petite qu'une ROI trop grande car on perdrait du signal. Certains préconisent même de dessiner une ROI tout autour de l'animal [5]. Cependant, certains protocoles expérimentaux nécessitent de définir des ROIs autour de plusieurs tumeurs sur la même

face de l’animal.

Sim et al. [15] ont montré la validité de modéliser la réaction de bioluminescence par un modèle pharmaco-cinétique à plusieurs compartiments. Afin d’isoler différentes tumeurs dans une image de BLI, notre hypothèse est que chacune constitue une source d’émission de photons dont la cinétique dépend de ses paramètres intrinsèques. Une méthode de séparation aveugle de sources semble pertinente dans ce cadre car le signal peut résulter d’un mélange inconnu des signaux sources. La Factorisation Non-Négative de Matrice (Non-Negative Matrix Factorization, NMF), introduite par [10], permet d’obtenir une déconvolution d’un signal de base et la séparation en une combinaison linéaire de  $K$  composants. La contrainte de positivité permet de donner des résultats naturellement interprétables dans de nombreux domaines.

Dans ce travail, nous étudions l’utilisation de la NMF comme méthode d’isolation des signaux sans a priori spatial. Pour évaluer cette méthode d’analyse nous avons utilisé un jeu d’images de BLI d’oncologie préclinique acquises sur des souris porteuses de deux tumeurs sur la face ventrale.

## 2 Matériels et méthodes

### 2.1 Jeux de données

Le jeu d’images utilisées est issu de six souris balbcJRJ ayant reçu par injection sous cutanée ventrale une suspension de cellules CT26-Luc ( $2.5 \times 10^5 / 50l$  milieu de culture). Les images BLI ont été acquises avec une caméra CCD (PhotonImager RT OPTIMA, BiospaceLab, Nesles la Vallée, France) en utilisant du sel de potassium de luciférine dilué dans du tampon phosphate 10mM injecté par voie i.p. à raison de 2 mg par souris anesthésiée, ce qui est en grand excès par rapport à la quantité de luciférase. Des acquisitions ont été faites tous les 2 jours à partir de 7 jours après l’implantation des tumeurs jusqu’au sacrifice de la souris dû à l’échec de leur traitement. En effet, chaque souris a été sacrifiée lorsque le volume de la tumeur, mesuré à l’aide d’un pied à coulisse, est devenu trop important (plus de  $1000 \text{ mm}^3$ ) ou lorsque le signal BLI est devenu trop fort (plus de  $5 \times 10^5 \text{ ph/s/cm}^2/\text{sr}$ ). L’acquisition de la luminescence a été initiée 8 min après l’injection du substrat pour une durée d’environ 7 min avec un échantillonnage temporel de 89ms par image et des pixels de taille  $0.12 \times 0.12 \text{ mm}^2$ .

Chaque acquisition BLI consiste en une série temporelle de signaux 2D acquis sur un champ de vision fixe. Une acquisition typique contient 4650 trames temporelles. Dans chaque trame, la valeur des pixels varie de 0 à 1000 et représente le nombre de photons captés.

Nous avons sélectionné uniquement les données provenant d’acquisitions où le volume est du même ordre de grandeur entre les 2 tumeurs dans le cadre d’une première approche et où il est supérieur à  $10 \text{ mm}^3$  pour avoir assez de signal. Cette sélection de données nous place dans un cadre idéal et sont établies dans le contexte d’une première analyse. Ces conditions devront être mises à l’épreuve dans de futurs travaux. Au total, avec 6 souris nous obtenons 15 images BLI et donc 30 tumeurs.

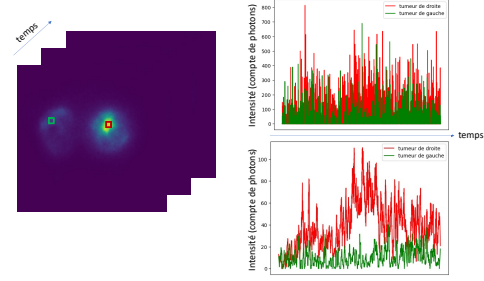


FIGURE 1 – Exemples de séries temporelles de BLI sur 2 pixels issus de 2 tumeurs, avant (en haut à droite) et après (en bas à droite) filtrage de Savitzky–Golay.

### 2.2 Algorithme

On note  $O(x, y, t)$  une image 2D + T de bioluminescence spatio-temporelle (pour une souris donnée et un jour donné) de taille  $M \times N \times T_{total}$  avec  $T_{total} = 4650$  trames temporelles qui représentent donc un comptage de photons. On note  $O_{ROI}(x, y, t)$  le signal restreint à une ROI autour des 2 tumeurs, obtenue manuellement.

La Figure 1 montre un exemple de séries temporelles sur deux pixels de tumeurs, avant et après filtrage de Savitzky–Golay [14] (pour améliorer la visualisation).

#### 2.2.1 Pré-traitement

Le signal est ensuite échantillonné avec une période d’échantillonnage notée  $T_e$  afin de réduire le temps de calcul puis lissé spatio-temporellement avec un filtre médian de taille  $3 \times 3 \times 3$ . On obtient finalement  $\hat{O}_{ROI}(x, y, t)$  de taille  $M \times N \times T$  où  $T = T_{total}/T_e$ . Enfin, le signal est rasterisé et concaténé dans une matrice 1D+T de taille  $P \times T$  où  $P = M \times N$  notée  $S(p, t)$ .

#### 2.2.2 Analyse factorielle par NMF

La NMF factorise une matrice positive  $S$  de taille  $P \times T$  en un produit de 2 matrices positives  $W$  et  $H$  ( $S \approx WH$ ), respectivement de tailles  $P \times K$  et  $K \times T$  représentant respectivement les charges et facteurs de la décomposition. Le seul hyperparamètre à régler est le nombre de facteurs  $K$ .

Les matrices de décomposition sont calculées en minimisant l’erreur (par exemple via la norme de Frobenius) entre la matrice d’origine et le résultat de la factorisation :

$$\min_{H \geq 0, W \geq 0} \|S - WH\|_F$$

Ce problème d’optimisation sous contrainte est résolu grâce à la méthode des poids multiplicatifs [9] dans lequel  $H$  et  $W$  sont mises à jour alternativement de manière itérative.

Dans notre contexte, nous faisons l’hypothèse que les profils temporels présents dans la matrice  $H$  correspondent aux dynamiques de chaque tumeur. On fixe donc  $K$  égal au nombre de tumeurs présentes dans le champ de vue.

Suivant cette hypothèse, on peut supposer obtenir une approximation  $\hat{S}_i(t)$  du signal  $S_i(t)$  correspondant à la tumeur  $i$  avec la méthode suivante. Notons  $W_i$  et  $H_i$  les facteurs appris avec la NMF associés à la tumeur  $i$ .  $W_i H_i$  est une matrice de

taille  $P \times T$  correspondant au signal reconstruit associé à la composante  $i$ . On définit donc  $\hat{S}_i(t) = (\sum_{p=1}^P W_i(p))H_i(t)$  comme la reconstruction sans a priori spatial du signal  $S_i(t)$ .

### 3 Résultats

Dans notre contexte, nous fixons  $K = 2$  pour identifier 2 tumeurs par image. La période d'échantillonnage  $T_e$  est réglée à  $T_e = 10$  trames. Dans la suite, on note donc  $W = (W_1, W_2)$  et  $H = (H_1, H_2)^T$ . Les décompositions sont initialisées grâce à l'algorithme NNDSVD [4] et le nombre maximal d'itérations est fixé à 200.

#### 3.1 Illustrations visuelles de la décomposition NMF

On illustre sur la Figure 2 le résultat de la NMF sur une image BLI, avant et après pré-traitement. Les cartes paramétriques des charges sont codées sur 2 canaux où le premier canal (en rouge) est  $W_1$  et le deuxième canal (en vert) est  $W_2$ . Pour aider à la visualisation du signal original, nous illustrons également les images d'accumulation où la valeur de chaque pixel correspond au nombre total de photons captés par ce pixel pendant toute la durée de l'acquisition.

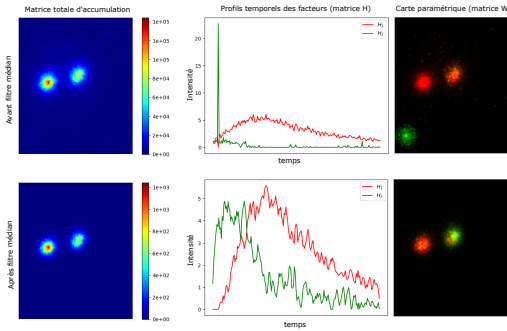


FIGURE 2 – Décomposition d'une image BLI avec la NMF : (haut) avant et (bas) après pré-traitement. Gauche : Image BLI d'accumulation ; Milieu : profils temporels des composantes  $H_1$  en rouge, et  $H_2$  en vert ; Droite : cartes paramétriques des charges  $W_i$  associées.

Dans cette figure, on observe tout d'abord que le filtrage permet de réduire l'effet perturbateur du bruit cosmique [1] sur la composante verte. La carte paramétrique issue de l'image BLI pré-filtrée confirme bien que la tumeur de gauche est associée à la première composante ( $H_1$ ) et que la tumeur de droite est associée à la deuxième ( $H_2$ ).

#### 3.2 Qualité de factorisation

Nous avons comparé les signaux reconstruits aux signaux originaux, au niveau de chaque tumeur avec des ROIs de référence manuellement délimitées.

Chaque signal reconstruit  $\hat{S}_i(t)$  est alors comparé au signal original  $S_i(t)$ , comme illustré sur la Figure 3. La métrique utilisée est la corrélation de Pearson  $r$ .

La NMF est comparée à un clustering K-Means sur les séries temporelles avec 3 classes : une pour chaque tumeur et une

pour le fond.

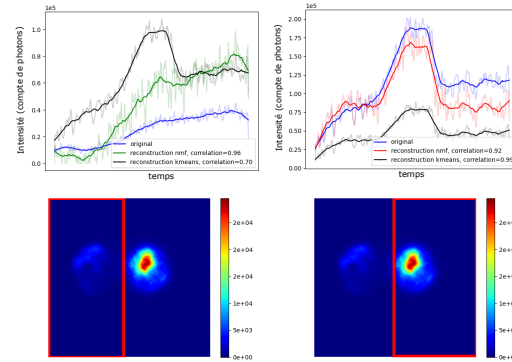


FIGURE 3 – Comparaison entre signaux de ROIs de 2 tumeurs (gauche et droite) reconstruits par NMF et K-Means et le signal original. Haut : signaux originaux et reconstruits ; Bas : ROI indiquée sur image d'accumulation.

Pour avoir une idée du caractère discriminant de notre méthode, nous avons comparé de la même manière pour chaque tumeur le signal  $S_i(t)$  avec  $\hat{S}_j(t)$  où  $j$  est l'autre tumeur présente dans la même image. Ce qui nous intéresse ici est de reconstruire le pattern du signal  $S_i(t)$ . Cependant, les signaux temporels que nous manipulons ici présentent de légères perturbations dues au bruit que nous ne souhaitons pas détecter mais qui dégradent la métrique de corrélation. Ces signaux sont donc lissés grâce à un filtrage de Savitzky–Golay avec une taille de fenêtre  $T_f$ , de réponse impulsionnelle  $h_{T_f}$ , est appliqué avant évaluation. Dans toute la suite, nous avons choisi empiriquement  $T_f = 31$  car c'est une fenêtre assez grande comparée à la taille des signaux temporels après échantillonnage ( $T = 465$ ) et cela permet d'atténuer les variations à faible fréquence). On note ainsi formellement pour chaque tumeur :

$$\Delta = r(h_{T_f} * S_i, h_{T_f} * \hat{S}_i) \text{ et } \delta = r(h_{T_f} * S_i, h_{T_f} * \hat{S}_j).$$

Les résultats sont synthétisés par la Figure 4. Selon le diagramme en boîtes, la méthode NMF est plus discriminante entre tumeurs (similarités  $\Delta > \delta$ ) que le K-means. En effet, la NMF donne  $\Delta = 0.76 \pm 0.2$  et  $\delta = 0.19 \pm 0.5$  alors que avec le K-means on a  $\Delta = 0.69 \pm 0.3$  et  $\delta = 0.49 \pm 0.5$ . Cependant ces 2 méthodes génèrent des résultats pour lesquels  $\Delta < \delta$ . Cela se produit 6 fois avec la NMF, lorsque le signal émis par une des tumeurs est très faible. La cause d'un tel phénomène peut être le traitement de la souris qui ferait artificiellement baisser le signal BLI mais pas le volume tumoral. La détection de telles réactions pourrait représenter un axe de travail intéressant pour les biologistes. Cela se produit avec le K-Means 4 fois de manière moins prévisible avec des cas où  $\Delta$  et  $\delta$  sont en fait très proches à cause du manque de discrimination de cette méthode.

#### 3.3 Résultats sur cinétique longue

Nous avons testé notre approche de décomposition sur une souris porteuse de 2 tumeurs et pour laquelle nous disposons d'une acquisition longue. L'algorithme a en fait été appliqué une fois sur la cinétique entière et une fois sur une fenêtre de 7 minutes au bout de 8 minutes après le début de l'acquisition

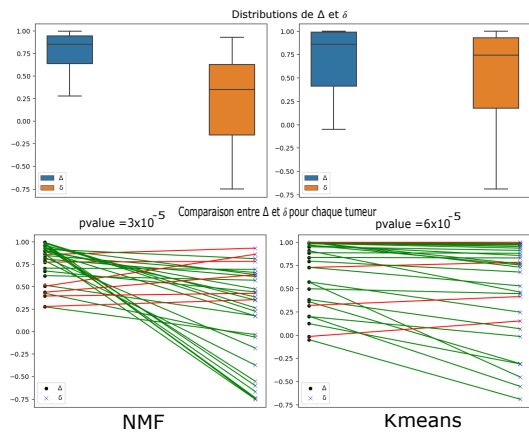


FIGURE 4 – Corrélations entre reconstructions de signaux de tumeurs avec la NMF (gauche) et le K-means (droite). Haut : Distributions des corrélations  $\Delta$  et  $\delta$ ; Bas : Paires de valeurs ( $\Delta$ ,  $\delta$ ) pour chaque tumeur - en rouge si  $\Delta < \delta$ . Les p-values correspondent à un test des rangs signés de Wilcoxon avec comme hypothèse nulle  $\delta > \Delta$ .

pour se remettre dans les conditions précédentes.

La reconstruction par NMF donne des corrélations de 0.88 et 0.82 sur la cinétique longue et de 0.64 et 0.51 pour la cinétique courte.

Ce résultat suggère qu'augmenter la durée de l'acquisition pourrait renforcer la robustesse de la décomposition.

## 4 Discussion et Conclusion

Dans ce travail, nous avons étudié l'utilisation de la NMF comme méthode de séparation aveugle de sources sans a priori spatial appliquée à des images de bioluminescence de suivi longitudinal de souris porteuses de 2 tumeurs. Les résultats semblent montrer que cet algorithme permet une factorisation du signal de bioluminescence plus performante qu'un algorithme de K-Means.

Cet algorithme a été étudié sur une cohorte d'acquisitions de cinétiques courtes et sur une unique acquisition de cinétique longue, sur lequel il semble performant. L'exploitation de cinétiques longues permettrait de tester la capacité de notre décomposition à récupérer les paramètres pharmaco-cinétiques de chaque tumeur.

Des travaux complémentaires sont en cours pour améliorer le modèle NMF et affiner nos analyses. Notamment, Bodvarsson et al. [3] propose de rajouter un facteur d'échelle aux composantes apprises par la NMF afin d'améliorer la quantification. L'algorithme FAMIS [2] pourrait également être étudié. Enfin, des contraintes spatiales pourraient aussi être ajoutées lors de la résolution du problème d'optimisation de la NMF.

## 5 Remerciements

Les auteurs tiennent à remercier la plateforme d'imagerie optique animale LIOPA, Faculté de Pharmacie (UMS US25 Inserm - UMS 3612 CNRS) Plateforme Imageries du Vivant de l'Université Paris

Cité. Ce travail a été financé en partie par des subventions des programmes France-BioImaging (ANR-10-INBS-04), PIA INCEPTION (ANR-16-CONV-0005 et Labex IBEID (ANR-10-LABX-62-IBEID). E.D. est financé par une bourse du programme PRPhD de la Région Ile-de-France. Les études ont été menées selon les recommandations de la Convention européenne sur la protection des animaux vertébrés et ont été validé par le Comité d'Ethique pour l'expérimentation animale de l'Université de Paris-Cité. Ces expériences soumises au Ministère de l'Enseignement Supérieur de la Recherche et de l'Innovation, sont référencées sous le numéro APAFIS #30778.

## Références

- [1] S. Akkoul, R. Lédée, R. Leconge, and R. Harba. A new adaptive switching median filter. *IEEE Signal Processing Letters*, 17(6) :587–590, 2010.
- [2] H. Benali, I. Buvat, F. Frouin, J. P. Bazin, and R. Di Paola. Foundations of factor analysis of medical image sequences : a unified approach and some practical implications. *Image and Vision Computing*, 12(6) :375–385, 1994.
- [3] B. Bodvarsson, L. K. Hansen, C. Svarer, and G. Knudsen. NMF on positron emission tomography. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 1, 2007.
- [4] C. Boutsidis and E. Gallopoulos. SVD based initialization : A head start for nonnegative matrix factorization. *Pattern Recognition*, 41(4) :1350–1362, apr 2008.
- [5] J. Cosette, R. B. Abdelwahed, S. Donnou-Triffault, C. Sautès-Fridman, P. Flaud, and S. Fisson. Bioluminescence-based tumor quantification method for monitoring tumor progression and treatment effects in mouse lymphoma models. *Journal of Visualized Experiments*, 2016(113), 2016.
- [6] K. Cui, X. Xu, H. Zhao, and S. T. Wong. A quantitative study of factors affecting in vivo bioluminescence imaging. *Luminescence*, 23(5) :292–295, 2008.
- [7] M. Keyaerts, J. Verschueren, T. J. Bos, L. O. Tchouate-Gainkam, C. Peleman, K. Breckpot, C. Vanhove, V. Caveliers, A. Bossuyt, and T. Lahoutte. Dynamic bioluminescence imaging for quantitative tumour burden assessment using IV or IP administration of d-luciferin : Effect on intensity, time kinetics and repeatability of photon emission. *European Journal of Nuclear Medicine and Molecular Imaging*, 35(5) :999–1007, 2008.
- [8] D. Lambrechts, M. Roeyffers, K. Goossens, J. Hofkens, T. Van De Putte, J. Schrooten, and H. Van Oosterwyck. A causal relation between bioluminescence and oxygen to quantify the cell niche. *PLoS ONE*, 9(5), 2014.
- [9] D. D. Lee and H. S. Seung. Algorithms for Non-negative Matrix Factorization. In *Advances in Neural Information Processing Systems*, 2000.
- [10] P. Paatero and U. Tapper. Positive matrix factorization : A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5(2) :111–126, 1994.
- [11] K. M. Parkins, V. P. Dubois, A. M. Hamilton, A. V. Makela, J. A. Ronald, and P. J. Foster. Multimodality cellular and molecular imaging of concomitant tumour enhancement in a syngeneic mouse model of breast cancer metastasis. *Scientific Reports*, 8(1), 2018.
- [12] B. W. Rice, M. D. Cable, and M. B. Nelson. In vivo imaging of light-emitting probes. *Journal of Biomedical Optics*, 6(4) :432, 2001.
- [13] R. T. Sadikot and T. S. Blackwell. Bioluminescence imaging. *Proc. Am. Thorac. Soc.*, 2(6) :537–40, 511–2, 2005.
- [14] A. Savitzky and M. J. E. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8) :1627–1639, 1964.
- [15] H. Sim, K. Bibee, S. Wickline, and D. Sept. Pharmacokinetic modeling of tumor bioluminescence implicates efflux, and not influx, as the bigger hurdle in cancer drug therapy. *Cancer Research*, 71(3) :686–692, 2011.
- [16] Y. Zilberman, Y. Gafni, G. Pelled, Z. Gazit, and D. Gazit. Bioluminescent imaging in bone. *Methods Mol. Biol.*, 455 :261–272, 2008.