

Форматы ATF-файлов

Содержание

1 GenePix Results (.GPR)	1
2 GenePix Array List (.GAL)	5
3 Spot (.Spot)	6

1 GenePix Results (.GPR)

GPR-файл описывает результат гибридизации микроматрицы и представляет из себя текстовый файл, состоящий из двух частей – заголовка и таблицы с данными. Файл соответствует спецификации ATF (Axon Text File), это значит, что он содержит только текстовые данные, разделённые табуляцией. ATF-файлы поддерживаются большинством табличных процессоров, например Microsoft Excel.

Первым в GPR-файле идёт заголовок, содержащий размеры пикселя отсканированного изображения, длины волн лазерного излучения, температуру сканера, мощность лазера и чувствительность фотоумножителя при сканировании, метод и коэффициенты нормализации (если она производилась при анализе изображения) и так далее. Далее следует таблица с данными о гибридизации по каждому споту (изображению сгибридизировавшего зонда). Столбцы этой таблицы описывают определённые характеристики спотов: положение на микроматрице (сектор, строка, столбец, ху-координаты); имя и описание гена, относящегося к зонду, представленном на изображении данным спотом; диаметр точки; средняя и медианная интенсивность точки для обеих длин волн; средняя и медианная фоновая интенсивности (фоновый шум) также для обеих длин волн; отношение интенсивностей (как средних, так и медианных) с вычетом соответствующих фоновых интенсивностей; столбец Flags, характеризующий качество спота, а также другие данные. Строки таблицы соответствуют спотам изображения микроматрицы.

Построчное описание заголовка GPR-файла:

ATF 1.0	Формат и версия файла.
29 48	Первое число – количество записей в заголовке, второе – количество столбцов в таблице с данными, которая следует за заголовком.
"Type=GenePix Results 3"	Тип ATF-файла.
"DateTime=2002/02/09 17:15:48"	Дата и время получения изображения.
"Settings=C:\Genepix\Genepix.gps"	Файл настроек, использованных при анализе.
"GalFile=C:\Genepix\Demo.gal"	GAL-файл с описанием структуры микроматрицы.
"PixelSize=10"	Размер пикселя изображения в мкм.
"Wavelengths=635 532"	Длины волн лазерного излучения в нм.
"ImageFiles=C:\Genepix\demo.tif 0 C:\Genepix\GenePix.tif 1"	Имя и путь соответствующих TIFF-файлов. TIFF-файл может включать в себя несколько изображений (мультистраничный TIFF-файл). Параметры 0 и 1 показывают номера каких изображений (страниц) в TIFF-файле следует использовать. Первым идёт изображение по первому каналу: 0 изображение из файла demo.tif. Вторым – по второму каналу: 1 изображение из файла GenePix.tif.
"NormalizationMethod=None"	Метод нормировки, если использовалась.

"NormalizationFactors=1 1"	Коэффициенты нормировки по каждому каналу.
"JpegImage=C:\Genepix\demo.jpg"	Имя и путь соответствующих JPEG-файлов.
"StdDev=Type 1"	Метод расчёта стандартного отклонения согласно настройкам опций.
"RatioFormulation=W1/W2 (635/532)"	Формула для расчёта отношения изображений, показывающая данные какого из изображений подставляются в числитель, а какого – в знаменатель.
"Barcode=00331"	Штрихкод, прочитанный с изображения.
"BackgroundSubtraction=LocalFeature"	Метод вычитания фоновой интенсивности согласно настройкам опций.
"ImageOrigin=0, 0"	Начало координат изображения.
"JpegOrigin=390, 4320"	Начало координат результирующего JPEG изображения.
"Creator=GenePix 4.1.1.4"	Версия ПО GenePix Pro, использованного при создании файла результатов.
"Scanner=GenePix 4000B [серийный номер]"	Тип и серийный номер сканера.
"FocusPosition=0"	Положение фокуса в мкм при сканировании изображения.
"Temperature=19.6127"	Температура сканера в градусах Цельсия.
"LinesAveraged=1"	Количество проходов каждой строки пикселей (от 1 до 3). Используется при малых интенсивностях свечения для уменьшения шума детектора.
"Comment=hyb 2673"	Комментарий пользователя к файлу.
"PMTGain=500 600"	Чувствительность фотоумножителя при сканировании.
"ScanPower=100 100"	Мощность лазера в процентах от максимальной для каждого канала.
"LaserPower=1 1"	Питание каждого лазера в вольтах.
"LaserOnTime=5 5"	Время лазера во включенном состоянии в минутах при сканировании по каждому каналу.
"Filters=<Empty> <Empty> "	Эмиссионные фильтры, использованные при сканировании.
"ScanRegion=100,100,2000,2000"	Координаты области сканирования в пикселях.
"Supplier=Производитель"	Поле из заголовка GAL-файла с именем производителя микроматрицы (см. 2).

Данные GPR-файла следуют за заголовком. Они упорядочены в таблицу: строки соответствуют спотам изображения микроматрицы, а столбцы – различным характеристикам этих спотов. Описание столбцов таблицы данных:

"Block"	Номер блока, на котором находится спот. Блоки микроматрицы нумеруются слева-направо, снизу-вверх (см. 2).
"Column"	Номер столбца в блоке, где находится спот.
"Row"	Номер строки в блоке, где находится спот.
"Name"	Имя зонда, изображением которого является данный спот. Не более 40 символов в GenePix Pro 4.0 и более ранних версиях и без ограничений начиная с GenePix Pro 4.1.
"ID"	Уникальный идентификатор зонда (краткое имя). Не более 40 символов в GenePix Pro 4.0 и более ранних версиях и без ограничений начиная с GenePix Pro 4.1.
"X" "Y"	XУ-координаты центра спота в мкм. Начало координат – левый верхний угол изображения.
"Dia."	Диаметр спота в мкм.
"F635 Median"	Медианная интенсивность пикселей спота на длине волны лазера 635 нм.

"F635 Mean"	Средняя интенсивность пикселей на длине волны лазера 635 нм.
"F635 SD"	Стандартное отклонение интенсивностей пикселей на длине волны лазера 635 нм.
"F635 CV"	Коэффициент вариации интенсивностей пикселей на длине волны лазера 635 нм.
"B635"	Фоновое значение интенсивности спота на длине волны лазера 635 нм, используемое GenePix Pro при вычислениях производных значений (в первых версиях GPR формата "B532" = "B635 Median", то есть при расчётах вычиталась медианная фоновая интенсивность). Этот столбец необходим, так как в GenePix Pro 5.0 есть разные методы вычитания фоновой интенсивности: глобальный и по отрицательным контрольным зондам. Если не выбран локальный метод ("BackgroundSubtraction≠LocalFeature"), то "B635" будет отличаться от "B635 Median".
"B635 Median"	Медианная фоновая интенсивность пикселей на длине волны лазера 635 нм.
"B635 Mean"	Средняя фоновая интенсивность пикселей на длине волны лазера 635 нм.
"B635 SD"	Стандартное отклонение фоновых интенсивностей пикселей на длине волны лазера 635 нм.
"B635 CV"	Коэффициент вариации фоновых интенсивностей пикселей на длине волны лазера 635 нм.
"% > B635+1SD"	Процент пикселей с интенсивностью более чем на одно стандартное отклонение большей фоновой интенсивности пикселей на длине волны лазера 635 нм.
"% > B635+2SD"	Процент пикселей с интенсивностью более чем на два стандартных отклонения большей фоновой интенсивности пикселей на длине волны лазера 635 нм.
"F635 % Sat."	Процент насыщенных пикселей (с максимальной интенсивностью) на длине волны лазера 635 нм.
"F635 Total Intensity"	Сумма интенсивностей всех пикселей спота на длине волны лазера 635 нм.
"SNR 635"	Соотношение сигнал-шум: $\frac{F635\ Mean - B635}{B635\ SD}$
"F635 Median - B635"	Медианная интенсивность спота на длине волны 635 нм с вычетом фоновой интенсивности.
"F635 Mean - B635"	Средняя интенсивность спота на длине волны 635 нм с вычетом фоновой интенсивности.

"F532 Median" "F532 Mean" "F532 SD" "F532 CV" "B532" "B532 Median" "B532 Mean" "B532 SD" "B532 CV" "% > B532+1SD" "% > B532+2SD" "F532 % Sat." "F532 Total Intensity" "SNR 532" "F532 Median - B532" "F532 Mean - B532"	Величины, аналогичные предыдущим, вычисленные для второго канала (длина волны лазера 532 нм).
"Ratio of Medians (635/532)"	Отношение медианных интенсивностей с вычетом фоновых интенсивностей: $\frac{F635 \text{ Median} - B635}{F532 \text{ Median} - B532}$
"Ratio of Means (635/532)"	Отношение средних интенсивностей с вычетом фоновых интенсивностей: $\frac{F635 \text{ Mean} - B635}{F532 \text{ Mean} - B532}$
"Median of Ratios (635/532)"	Медиана попиксельных отношений интенсивностей с вычетом фоновых интенсивностей: $\text{Median}_{k=1}^N \left\{ \frac{F635_k - B635}{F532_k - B532} \right\}$
"Mean of Ratios (635/532)"	Геометрическое среднее попиксельных отношений интенсивностей с вычетом фоновых интенсивностей: $\mu = \sqrt[N]{\prod_{k=1}^N \frac{F635_k - B635}{F532_k - B532}}$
"Ratios SD (635/532)"	Геометрическое стандартное отклонение отношений интенсивностей пикселей: $\text{Ratios SD} = \exp \left(\sqrt{\frac{\sum_{k=1}^N \left(\ln \left(\frac{F635_k - B635}{F532_k - B532} \right) - \ln \mu \right)^2}{N}} \right)$
"Rgn Ratio (635/532)"	Регрессионное отношение каждого пикселя в круглой области с центром, совпадающим с центром спота, и диаметром, равным удвоенному диаметру спота.
"Rgn R2 (635/532)"	Коэффициент определения для данного регрессионного значения.

"F Pixels"	Общее число пикселей.
"B Pixels"	Общее число фоновых пикселей.
"Circularity"	Циркулярность – степень близости формы спота к круглой: 100 – идеально круглая, 0 – абсолютно некруглая. Квадратные споты имеют циркулярность 79. Рассчитывается с использованием метрики, основанной на вариации расстояний от каждого граничного пикселя спота до центра масс этого спота.
"Sum of Medians (635/532)"	Сумма медианных интенсивностей с вычетом фоновых интенсивностей.
"Sum of Means (635/532)"	Сумма средних интенсивностей с вычетом фоновых интенсивностей.
"Log Ratio (635/532)"	Логарифм по основанию 2 от отношения медианных интенсивностей: $\log_2(\text{Ratio of Medians}(635/532)) = \log_2\left(\frac{F635 \text{ Median} - B635}{F532 \text{ Median} - B532}\right)$
"Flags"	Качество спота: 100 – идеально качественный спот, 0 – абсолютно некачественный спот.
"Normalize"	Статус нормировки спота: 1 – нормирован, 0 – ненормирован.
"Autoflag"	Был ли споту присвоен флаг из диалогового окна Flag Features. Применимо только к хорошим и плохим флагам.

2 GenePix Array List (.GAL)

GAL-файл описывает расположение зондов на микроматрице. Он очень похож на GPR-файл. В частности, GAL-файл также является разновидностью ATF-файла, и он тоже состоит из заголовка и таблицы с данными. Заголовок GAL-файла содержит общую информацию о микроматрице. После заголовка следует таблица с данными, описывающая размещение и название зондов микроматрицы.

Построчное описание заголовка GAL-файла. Обязательные поля указаны, все остальные поля необязательны:

ATF 1.0	Обязательное поле. Формат и версия файла.
8 5	Обязательное поле. Первое число – количество записей в заголовке, второе – количество столбцов в таблице с данными, которая следует за заголовком.
"Type=GenePix Array List v1.0"	Обязательное поле. Тип файла.
"BlockCount=4"	Число блоков, описанных в файле.
"BlockType=0"	Тип описанных блоков: 0 - прямоугольные, 1 - orange-packing #1, 2 - orange-packing #2.
"URL=..."	URL-ссылка, как правило, на базу данных в Интернет.
"Supplier=Производитель"	Компания-производитель.
"ArrayerSoftwareName=Printer Robot User Interface"	Приложение-споттер, под управлением которого были нанесены зонды на микроматрицу.
"ArrayerSoftwareVersion=1.1"	Версия приложения-споттера.
"ArrayName= MouseApoptosisProteins 4000"	Имя микроматрицы.
"ArrayRevision=2.7"	Версия микроматрицы.
"SlideBarcode=abc0011, abc0012, abc0013"	Штрихкоды поддерживаемые GAL-файлом, используемые для штрихкод-автоматизации.

"BlockN=xOrigin, yOrigin, xFeatures, xSpacing, yFeatures, ySpacing"	Описание блока с номером N. xOrigin и yOrigin – координаты левого верхнего угла блока в мкм (начало координат – левый верхний угол микроматрицы); xFeatures – число столбцов с зондами; xSpacing – расстояние между столбцами в мкм; yFeatures – число строк с зондами; ySpacing – расстояние между строками в мкм.
"Пользовательское поле"	Заголовочная запись пользователя. Передаётся на вход GenePix Pro и включается в GPR-файл результатов.

Блоки микроматрицы нумеруются слева-направо, снизу-вверх:

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

Данные GAL-файла следуют за заголовком. Они упорядочены в таблицу: строки соответствуют зондам микроматрицы, а столбцы описывают положение зондов и их название. Описание столбцов таблицы данных:

"Block"	Номер блока, на котором находится зонд. Обязательный столбец.
"Column"	Номер столбца в блоке, где находится зонд. Обязательный столбец.
"Row"	Номер строки в блоке, где находится зонд. Обязательный столбец.
"Name"	Имя зонда. Не более 40 символов в GenePix Pro 4.0 и более ранних версиях и без ограничений начиная с GenePix Pro 4.1. Обязательный столбец.
"ID"	Уникальный идентификатор зонда (краткое имя). Не более 40 символов в GenePix Pro 4.0 и более ранних версиях и без ограничений начиная с GenePix Pro 4.1. Обязательный столбец.
"User Defined"	Столбец с информацией пользователя.

3 Spot (.Spot)

Spot – это один из R-пакетов по обработке изображений микроматриц (доступен на www.hca-vision.com/product_spot). В этом пакете есть функция `spotresults`, которая проводит анализ изображения. Результат работы этой функции – таблица значений, строки которой соответствуют спотам изображения, а столбцы – характеристикам этих спотов. Данные могут сохранены в файл в текстовом виде с табуляционным разделением (чтобы соответствовать спецификации ATF). Такой Spot-файл не содержит заголовка, он содержит только таблицу данных. Описание столбцов таблицы данных:

indexs	Индекс (номер) спота.
grid_r	Индекс строки блока (левый верхний угол), в котором находится спот.
grid_c	Индекс столбца блока (левый верхний угол), в котором находится спот.
spot_r	Индекс строки спота в блоке, в котором находится этот спот.
spot_c	Индекс столбца спота в блоке, в котором находится этот спот.
area	Пиксельная площадь спота – число пикселей в сегментированном споте.
Gmean	Средняя интенсивность пикселей спота по зелёному каналу.
Gmedian	Медианная интенсивность пикселей спота по зелёному каналу.
GIQR	Межквартильный диапазон логарифмов от интенсивностей пикселей спота по зелёному каналу.
bgGmean	Средняя фоновая интенсивность пикселей по зелёному каналу.
bgGmedian	Медианная фоновая интенсивность пикселей по зелёному каналу.

bgGSD	Межквартильный диапазон логарифмов от фоновых интенсивностей пикселей по зелёному каналу.
valleyG	Медиана среди медианных значений интенсивностей пикселей в 4 соседних “долинах” по зелёному каналу. “Долина” – это квадратная область между спотами.
Rmean Rmedian RIQR bgRmean bgRmedian bgRSD valleyR	Характеристики, аналогичные предыдущим, для красного канала.
morphG	Величина морфологического открытия (morphological opening) по зелёному каналу в номинальном центре спота. При расчёте открытий и эрозий используется прямоугольный структурный элемент (structuring element) с длинами сторон, равными удвоенным средним межспотовым расстояниям по горизонтали и вертикали.
morphG.erode	Величина морфологической эрозии (morphological erosion) по зелёному каналу в номинальном центре спота.
morphG.close.open	Величина морфологического закрытия (morphological closing), сопровождаемого морфологическим открытием, по зелёному каналу в номинальном центре спота.
morphR morphR.erode morphR.close.open	Характеристики, аналогичные предыдущим, для красного канала.
logratio	Логарифм по основанию 2 от отношения интенсивностей $\frac{R}{G}$, скорректированный с учётом фоновой интенсивности с использованием $\frac{R}{G_{median}}$ для значения интенсивности спота и $\frac{morphR}{G}$ для фоновой интенсивности.
perimeter	Периметр спота – количество граничных пикселей спота, т.е. таких, которые соседствуют с каким-либо пикселем, находящимся за пределами спота. При этом рассматриваются диагональные соседи, т.е. пиксель может иметь до 8 соседей.
circularity	Циркулярность – степень близости формы спота к круглой: $\frac{area \cdot 4\pi}{perimeter^2}$. Равна 1 для идеально круглых спотов и больше 1 для всех других форм.
badspot	Оценка качества спота – флаг, равный 1, если $area > \bar{H} \cdot \bar{V}$, где \bar{H} и \bar{V} – средние межспотовые расстояния по горизонтали и вертикали, и равный 0 в противном случае.