

В геноме человека пересчитали число генов

Опубликована новая оценка числа генов в геноме человека. При этом исследователи учли как гены, кодирующие белки, так и гены, не кодирующие белки, а синтезирующие РНК с другими функциями. Итого, всего у человека насчитали 43 162 гена, из них 21 306 кодирующих и 21 856 не кодирующих.

В статье, [появившейся на сайте препринтов](#), опубликован новый подсчет числа генов в геноме человека. Как это ни покажется странным, хотя геном человека был прочитан в 2001 году, определить точное число генов в нем – не такая простая задача. Раньше полагали, что у человека от 50 до 100 тысяч генов. Секвенирование генома сократило это число до 30-40 тысяч. Затем, по мере функционального исследования генома, ученые снизили число генов до 20-25 тысяч, эта оценка была сделана в 2004 году. Проведенный в 2007 году сравнительный геномный анализ указал на число 20 500 генов. В 2010 году число генов было оценено в 22 500. Последние оценки, относящиеся в 2017 году: 20 054 генов (в базе NCBI's RefSeq) и 19 817 генов (в каталоге Gencode).

Авторы нового исследования провели свою оценку числа генов. И, прежде всего, поставили принципиальный вопрос, что считать геном. Со времени проекта «Геном человека» принято считать таковыми только гены, кодирующие белки. Но теперь открыто множество участков ДНК, с которых синтезируется РНК, не несущая информацию о строении молекулы белка, а выполняющая иные функции. В новом пересчете они тоже именуется генами, не кодирующими белки генами.

Работа основана на анализе данных по 9795 секвенированным геномам человека из клеток, относящихся к 31 ткани сотен человек (проект GTEx). По оценкам авторов в геноме человека всего насчитывается 43 162 гена, из которых 21 306 кодируют белки и 21 856 – не кодируют. Ученые подсчитали также число транскриптов (то есть синтезируемых с генов РНК) и получили оценку 323 824 транскрипта, в среднем по 7,5 на один ген. По сравнению с последним каталогом генов человека, они нашли 4 998 новых генов (1 178 кодирующих и 3 819 не кодирующих). Описали также 97 511 новых вариантов альтернативного сплайсинга (получения с одного гена разных белков путем разной комбинации экзонов). А также обнаружили свыше 30 миллионов дополнительных транскриптов с более чем 650 тысяч сайтов, почти все из которых были нефункциональными и представляют собой «транскрипционный шум» в клетке.

Источник:

Thousands of large-scale RNA sequencing experiments yield a comprehensive new human gene list and reveal extensive transcriptional noise

Mihaela Pertea et al.

<https://www.biorxiv.org/content/early/2018/05/29/332825>