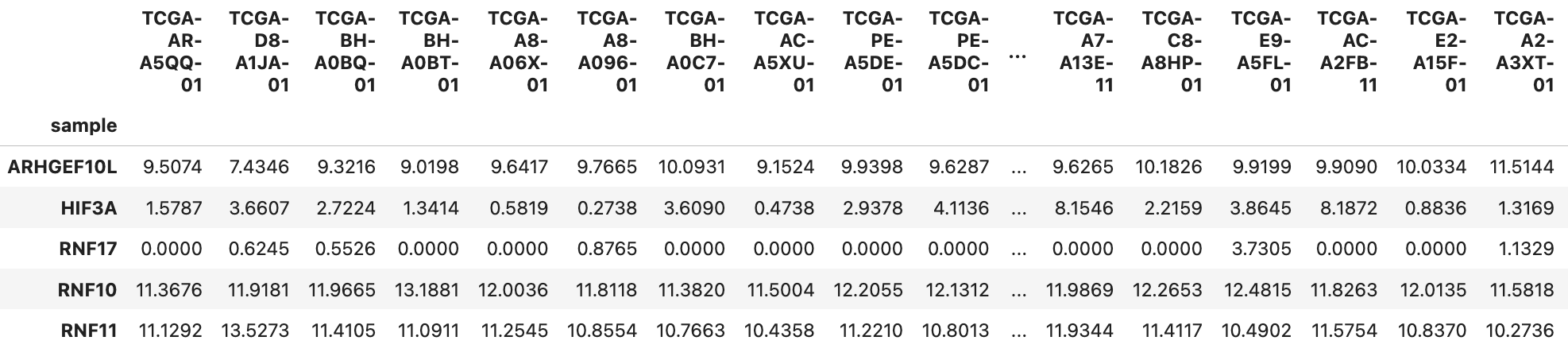
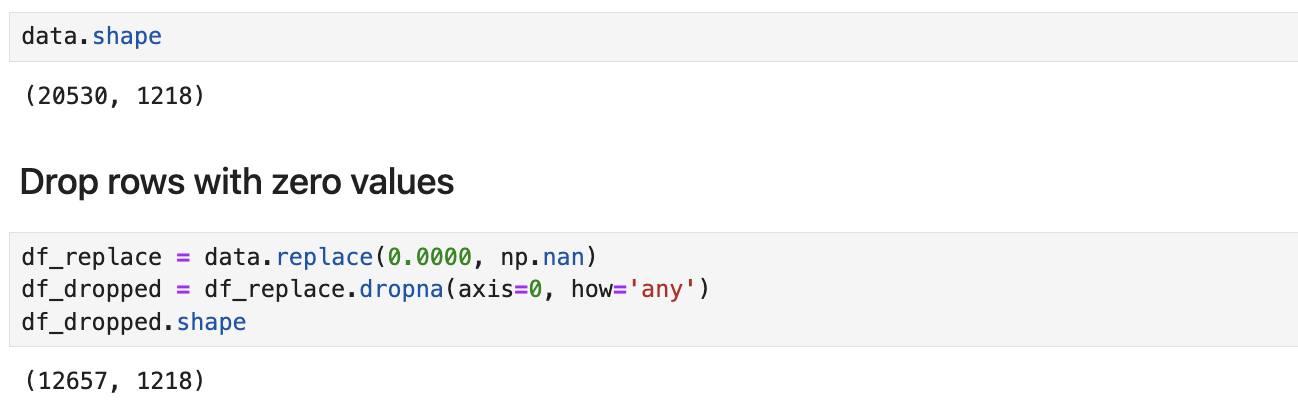
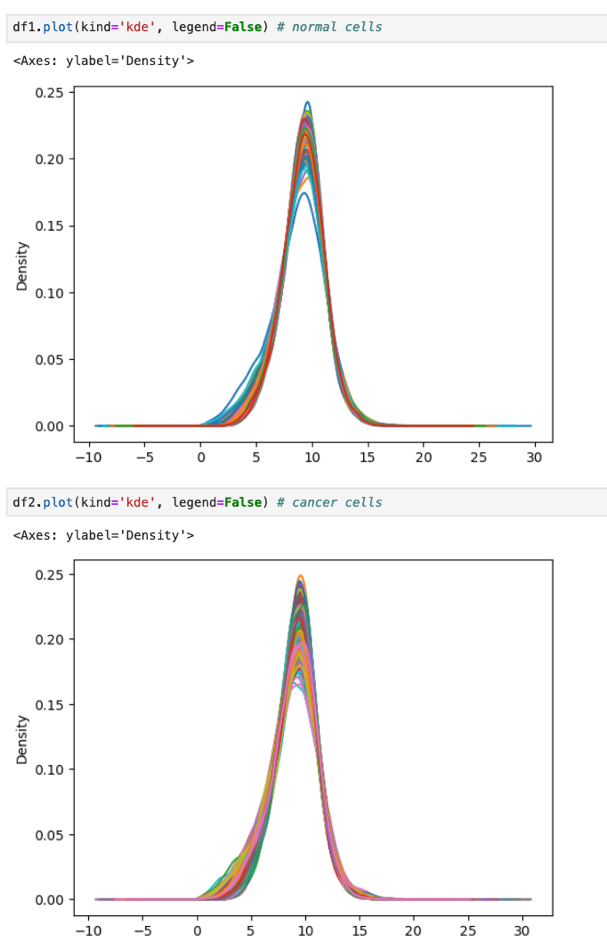
**მონაცემთა მომზადება**

მონაცემთა ბაზა შეგიძლიათ იხილოთ შემდეგ მისამართზე: <https://tcga-xena-hub.s3.us-east-1.amazonaws.com/download/TCGA.BRCA.sampleMap%2FHiSeqV2.gz>

მონაცემები ჩატვირთულია Jupyter Notebook-ში, და ქვემოთ ნაჩვენებია მისი ვიზუალიზაცია.

როგორც შეინიშნება, ბაზა შეიცავს მონაცემებს, რომლებიც ანალიზისთვის გამოუსადეგარია, ამიტომ პირველ რიგში გავფილტრეთ მონაცემები. ყველა ნულოვანი მნიშვნელობა გადავაქციეთ NaN-ად და ყველა სტრიქონი, რომელიც შეიცავდა NaN-ს, ამოვშალეთ. შედეგად, სტრიქონების რაოდენობა 20530-დან 12657-მდე შემცირდა.

ამის შემდეგ, მონაცემები გავყავით ორ DataFrame-ად: ერთში კიბოს უჯრედების გენების ექსპრესიის მნიშვნელობები და მეორეში ნორმალური უჯრედების გენების ექსპრესიის მნიშვნელობები. ნიმუშები, რომლებიც მთავრდება 11-ით, მიუთითებდნენ ჯანმრთელ უჯრედებზე, ხოლო 01-ით დამთავრებულები — კიბოთი დაავადებულ უჯრედებზე.

როგორც ჩანს, ორივე DataFrame-ში მონაცემები ნორმალურად არის განაწილებული.

**მონაცემთა ანალიზი**

Differential expression analysis ანალიზი წარმოადგენს მეთოდს, რომელიც გამოიყენება გენებისა და სხვა ბიოლოგიური მოლეკულების ექსპრესიის განსხვავებების შესასწავლად სხვადასხვა პირობებში (მაგალითად, ნორმალურ და დაავადებულ უჯრედებში). Limma არის ერთ-ერთი პოპულარული პაკეტი, რომელიც ფართოდ გამოიყენება ამ ტიპის ანალიზისთვის. Limma მეთოდის გამოყენების შემდეგ შესაძლებელია მნიშვნელოვანი განსხვავებით გამოხატული გენების იდენტიფიცირება, რომლებიც დაკავშირებულია ბიოლოგიურ განსხვავებებთან, როგორიცაა დაავადების სპეციფიკური მარკერების იდენტიფიცირება ან ახალი თერაპიული მიზნების აღმოჩენა. Limma მეთოდი ცნობილია თავისი ეფექტურობითა და სიზუსტით, რაც მას მნიშვნელოვან ინსტრუმენტად აქცევს ბიოინფორმატიკაში.

R-ის limma პაკეტი გამოვიყენეთ Python-ში rpy2 ბიბლიოთეკის მეშვეობით. პირველ რიგში, შევქმენით გაერთიანებული DataFrame (df1 და df2) და ჯგუფის ლეიბლები, რომლებიც აღწერს თითოეული სვეტის მდგომარეობას (ნორმალური ან კიბო). ვინაიდან მონაცემები წინასწარ დამუშავდა (ნორმალიზებულია და დაბალი გამოხატვის გენები ამოშლილია), ისინი მზადაა Limma-ს შემდგომი ანალიზისთვის. შემდეგი ეტაპებია:

**დიზაინი მატრიცის შექმნა:** გათვალისწინებული პირობების (მაგალითად, ნორმალური და დაავადებული უჯრედები) შესაბამისი დიზაინი მატრიცის შექმნა, რომელიც მოიცავს ჯგუფების ლეიბლებს.

**კონტრასტ მატრიცის შექმნა:** ქმნის კონტრასტ მატრიცას, რომელიც გამოიყენება კიბოსა და ნორმალური გენების გამოხატვის შედარებისთვის.

**ევალუაცია:** მეთოდი იყენებს სტატისტიკურ ტესტებს გენის გამოხატვის დონეების შესადარებლად სხვადასხვა პირობებში; მნიშვნელოვანი განსხვავებით გამოხატული გენების სიების შექმნა.