



ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Τίτλος Πτυχιακής: Πρόσφατες τεχνολογίες single cell και μεθοδολογίες ανάλυσης
τους

Υποψήφιος : Κωνσταντόπουλος Αντώνιος Π2017116

Επιβλέπων : Αριστείδης Βραχάτης

ΚΕΡΚΥΡΑ ,2024

ΕΥΧΑΡΙΣΤΙΕΣ

Με την ολοκλήρωση της πτυχιακής μου εργασίας, θα ήθελα να εκφράσω την ευγνωμοσύνη μου σε όλους εκείνους που συνέβαλαν στην πραγματοποίησή της.

Πρώτα απ' όλα, θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου, κ. Αριστείδη Βραχάτη, για την πολύτιμη καθοδήγηση, την υπομονή και την αμέριστη υποστήριξή του καθ' όλη τη διάρκεια της ερευνητικής διαδικασίας. Οι γνώσεις και οι συμβουλές του ήταν ανεκτίμητες και η συμβολή του καθοριστική για την ολοκλήρωση αυτής της εργασίας.

Ιδιαίτερες ευχαριστίες οφείλω στην οικογένειά μου για την συνεχή στήριξη και την ενθάρρυνση που μου παρείχαν καθ' όλη τη διάρκεια των σπουδών μου. Η αγάπη και η πίστη τους στις δυνατότητές μου υπήρξαν πηγή δύναμης και έμπνευσης.

Επιπλέον, θα ήθελα να ευχαριστήσω τους φίλους και συμφοιτητές μου για την πολύτιμη συνεργασία, τη συντροφιά και την αλληλοϋποστήριξη που μοιραστήκαμε στα χρόνια των σπουδών μας.

Τέλος, ευχαριστώ όλους όσους συνέβαλαν άμεσα ή έμμεσα στην ολοκλήρωση αυτής της εργασίας και συνέβαλαν στην ακαδημαϊκή και προσωπική μου ανάπτυξη.

Με εκτίμηση,
Κωνσταντόπουλος Αντώνιος

Πίνακας Περιεχομένων

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ	5
ΚΕΦΑΛΑΙΟ 2: scRNA – seq	7
2.1. Εισαγωγή	7
2.2. Γενικές πληροφορίες και ροή εργασίας	9
2.3. Ποιοτικός έλεγχος	11
2.4. Κανονικοποίηση	12
2.5. Διόρθωση Παρτίδας	13
2.6. Καταλογισμός (Imputation)	15
2.7. Επιλογή χαρακτηριστικών και μείωση διαστάσεων	16
2.8. Οπτικοποίηση και ομαδοποίηση	17
2.9. Σχολιασμός συστάδας (Cluster Annotation)	19
2.10. Ανάλυση τροχιάς	19
2.11. Ανάλυση γονιδιακής έκφρασης	21
2.12. Ζητήματα πειραματικού σχεδιασμού	23
2.13. Εφαρμογές	24
ΚΕΦΑΛΑΙΟ 3: ΧΩΡΙΚΗ ΜΕΤΑΓΡΑΦΙΚΗ	27
3.1. Εισαγωγή	27
3.2. Χωρική μεταγραφική στην τρέχουσα βιοϊατρική έρευνα	29
3.3. Τεχνολογίες Χωρικής Μεταγραφικής	31
3.3.1. Τεχνολογίες που βασίζονται στην απεικόνιση	31
3.3.2. Τεχνολογίες που βασίζονται σε αλληλούχιση	34
3.4. Βήματα Χωρικής Μεταγραφικής	35
3.4.1. Προεπεξεργασία χωρικών μεταγραφικών Δεδομένων	35
3.4.2. Προσδιορισμός χωρικών χαρακτηριστικών	36
3.4.3. Αποσυνέλιξη	37
3.4.4. Καταλογισμός και χαρτογράφηση μεμονωμένων κυττάρων	38
3.4.5. Συμπερασματολογία αλληλεπίδρασης κυττάρου-κυττάρου	39
ΚΕΦΑΛΑΙΟ 4: ATAC-seq	41

4.1. Εισαγωγή	41
4.2. Αρχή και διαδικασίες της ATAC-seq	43
4.3. Πλεονεκτήματα και περιορισμοί	45
4.4. Βελτίωση της ATAC-seq	47
4.5. Εφαρμογές	48
4.5.1. Χαρτογράφηση του προσβάσιμου τοπίου χρωματίνης.....	48
4.5.2. ATAC-seq στην εμβρυϊκή ανάπτυξη	49
4.5.3. Ανίχνευση καρκίνου	50
ΚΕΦΑΛΑΙΟ 5: ΣΥΝΔΥΑΣΜΟΣ ΤΕΧΝΟΛΟΓΙΩΝ	52
5.1. Σημασία συνδυασμού τεχνολογιών.....	52
5.2. Παραδείγματα συνδυαστικής χρήσης τεχνολογιών	54
5.2.1. Συνδυαστική χρήση scRNA-seq και ATAC-seq στην περίπτωση της παιδικής λευχαιμίας 54	
5.2.2. Ολοκληρωμένη μονοκυτταρική ανάλυση RNA-Seq και ATAC-Seq επιθηλιακών κυττάρων κερατοειδούς ποντικού	56
5.2.3. Συνδυαστική χρήση scRNA-Seq και ATAC-Seq σε ασθενείς με αγκυλοποιητική σπονδυλίτιδα.....	57
5.3. Προκλήσεις.....	58
ΚΕΦΑΛΑΙΟ 6 : ΣΥΜΠΕΡΑΣΜΑΤΑ	60
ΒΙΒΛΙΟΓΡΑΦΙΑ – ΠΗΓΕΣ	62

Λίστα Εικόνων

Εικόνα 1: Βήματα scRNA-seq	10
Εικόνα 2: Είδη Χωρικής Μεταγραφικής.....	32
Εικόνα 3: Βασική μεθοδολογία ATAC - seq.....	42
Εικόνα 4: Βήματα ATAC - seq.....	43
Εικόνα 5: Παρουσίαση αποτελεσμάτων	56

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ

Η εξέλιξη των τεχνολογιών ανάλυσης single – cell έχει σημειώσει σημαντική πρόοδο για τις βιολογικές επιστήμες, προχωρώντας από τις πρώιμες τεχνικές μικροσκοπίας σε εξελιγμένες μεθοδολογίες υψηλής απόδοσης. Αρχικά, η ανάλυση single – cell περιοριζόταν σε μεγάλο βαθμό σε οπτικές παρατηρήσεις μέσω μικροσκοπίων, επιτρέποντας στους ερευνητές να εντοπίζουν τύπους κυττάρων και να παρατηρούν βασικές κυτταρικές διεργασίες. Ωστόσο, οι περιορισμοί στην ικανότητα επίλυσης και η αδυναμία ανίχνευσης μοριακών λεπτομερειών περιόρισαν την πειραματική έρευνα. Η εμφάνιση της κυτταρομετρίας ροής στη δεκαετία του 1970 σηματοδότησε μια σημαντική πρόοδο, επιτρέποντας την ποσοτική ανάλυση πολλαπλών παραμέτρων σε επίπεδο ενός κυττάρου. Αυτή η τεχνολογία χρησιμοποίησε φθοριστικά επισημασμένα αντισώματα για να αναλύσει και να ταξινομήσει τα κύτταρα με βάση συγκεκριμένους δείκτες, διευκολύνοντας μελέτες στην ανοσολογία, τον καρκίνο και την έρευνα βλαστοκυττάρων.

Ο 21ος αιώνας οδήγησε σε μια επανάσταση με την ανάπτυξη της τεχνολογίας scRNA-seq, η οποία παρείχε μια άνευ προηγουμένου ανάλυση των προφίλ γονιδιακής έκφρασης σε επίπεδο μεμονωμένων κυττάρων. Τεχνικές όπως η μικρορευστολογία και τα συστήματα με βάση τα σταγονίδια επέτρεψαν την ταυτόχρονη σύλληψη και αλληλουχία χιλιάδων κυττάρων, ενισχύοντας δραματικά την κλίμακα και την ακρίβεια των αναλύσεων ενός κυττάρου. Σε συνδυασμό με τις εξελίξεις στην Υπολογιστική Βιολογία, αυτές οι τεχνολογίες επέτρεψαν τη λεπτομερή χαρτογράφηση της κυτταρικής ετερογένειας, αποκαλύπτοντας νέες γνώσεις για πολύπλοκους ιστούς, αναπτυξιακές διαδικασίες και μηχανισμούς ασθενειών. Επιπλέον, καινοτομίες όπως οι διαταραχές που βασίζονται στο CRISPR και η χωρική μεταγραφική διεκπεραίωση το πεδίο των μελετών ενός κυττάρου, επιτρέποντας τη λειτουργική γονιδιωματική και τη χωρικά διαλυμένη μοριακή διαμόρφωση προφίλ, αντίστοιχα. Αυτές οι εξελίξεις συλλογικά ωθούν τα όρια της κατανόησης της κυτταρικής λειτουργίας και οργάνωσης.

Στα πλαίσια της συγκεκριμένης εργασίας, βασικός στόχος είναι η ανάλυση αυτών των τεχνολογιών, δίνοντας έμφαση στα βήματα που απαιτούνται για την ανάλυση των δεδομένων, καθώς και η παρουσίαση συνδυαστικών μεθόδων για την παροχή πιο εξειδικευμένης λειτουργικότητας.

Πιο συγκεκριμένα, τα βασικά σημεία της εργασίας είναι:

1. Ανάλυση της τεχνολογίας scRNA-seq
2. Ανάλυση της χωρικής μεταγραφικής.
3. Ανάλυση της ATAC-seq
4. Παρουσίαση συνδυαστικών προσεγγίσεων scRNA-seq και ATAC-seq

ΚΕΦΑΛΑΙΟ 2: scRNA – seq

2.1. Εισαγωγή

Για πολλά χρόνια οι ερευνητές προσπάθησαν να κατανοήσουν την πολυπλοκότητα των ιστών, των οργάνων και των οργανισμών. Για να αποκτήσουν αυτή την κατανόηση, πολλές μελέτες έχουν επικεντρωθεί στον χαρακτηρισμό των κυττάρων, επαναπροσδιορίζοντας το κύτταρο ως όχι μόνο τη δομική αλλά και τη λειτουργική μονάδα της ζωής.

Παραδοσιακά, τα κύτταρα έχουν ταξινομηθεί με βάση τη μορφολογία τους ή με βάση την έκφραση ορισμένων πρωτεϊνών σε λειτουργικά διακριτά περιβάλλοντα, αλλά η έλευση των τεχνικών NGS άνοιξε το δρόμο για την ανίχνευση και την ποσοτική ανάλυση του κυτταρικού περιεχομένου. Η υψηλή ποσότητα δεδομένων που παράγονται στα σύγχρονα πειράματα γονιδιωμικής και μεταγραφικής επέτρεψε τον καλύτερο χαρακτηρισμό της αρχιτεκτονικής των γονιδιωμάτων και της πολυπλοκότητας των μοριακών μηχανισμών που διέπουν την κυτταρική δραστηριότητα, επιτρέποντας μια όλο και πιο ακριβή και σε βάθος απεικόνιση της πλαστικότητας των κυττάρων σε δυναμικές διεργασίες όπως η ανάπτυξη, η διαφοροποίηση και η εξέλιξη της νόσου.

Η σύγχρονη γνώση της κυτταρικής και Μοριακής Βιολογίας προέρχεται σε μεγάλο βαθμό από πειράματα αλληλούχισης RNA (RNA-seq). Τα τελευταία 20 χρόνια, η ποσοτικοποίηση του μεταγραφικού συστήματος έχει διαμορφώσει την κατανόησή μας για τους μηχανισμούς που είναι υπεύθυνοι για φαινόμενα, όπως η εναλλαξιμότητα της διαδικασίας ματίσματος mRNA, η ρύθμιση της γονιδιακής έκφρασης με μη κωδικοποιητικά και ενισχυτικά RNA αντίστοιχα και η αντοχή στα φάρμακα σε ορισμένους τύπους καρκίνου, καθιστώντας μια κοινή και ισχυρή τεχνολογία κατάλληλη για βιοϊατρική έρευνα.

Η προσαρμογή και η εξέλιξη του RNA-seq καθοδηγήθηκε από τις τεχνολογικές εξελίξεις και οδήγησε σε προοδευτική αύξηση της ανάλυσης. Ξεκινώντας από το λεγόμενο "bulk" RNA-seq, ικανό να μετρήσει τα μέσα επίπεδα γονιδιακής έκφρασης συνόλων εκατομμυρίων κυττάρων, καταλήξαμε στην scRNA-seq [1] που, επιτρέποντας το προφίλ του μεταγραφώματος μεμονωμένων κυττάρων, αποκάλυψε σπάνιες κυτταρικές ιδιότητες και τη βιολογικά σημαντική

μεταβλητότητα κυττάρου προς κύτταρο, θέτοντας τις βάσεις για μελέτες προσανατολισμένες στην ετερογένεια.

Καθώς τα πειραματικά πρωτόκολλα έχουν βελτιωθεί γρήγορα, οι υπολογιστικές ροές εργασίας για την επεξεργασία των δεδομένων έχουν επίσης βελτιωθεί, λαμβάνοντας υπόψη την αυξημένη απόδοση των πειραμάτων scRNA. Ο τρέχων τυποποιημένος αγωγός ανάλυσης αποτελείται από δύο κύριες ενότητες: προεπεξεργασία, συμπεριλαμβανομένων όλων των απαραίτητων βημάτων για τον καθαρισμό της μήτρας δεδομένων από ανεπιθύμητες πηγές πληροφοριών (ποιοτικός έλεγχος, ομαλοποίηση, διόρθωση δεδομένων, επιλογή χαρακτηριστικών και μείωση διαστάσεων) και ανάλυση σε επίπεδο κυττάρων και γονιδίων, που χρησιμοποιείται για την εξαγωγή βιολογικών πληροφοριών και την περιγραφή του υποκείμενου βιολογικού συστήματος. Για κάθε ένα από αυτά τα βήματα, οι βιολόγοι ανέπτυξαν μια σειρά μεθόδων που αποδίδουν καλύτερα σε διαφορετικές εργασίες και παραμετροποιήσεις, καθιστώντας δύσκολη τη δημιουργία γενικεύσιμων ροών εργασίας για την ανάλυση των πειραμάτων ενός κυττάρου.

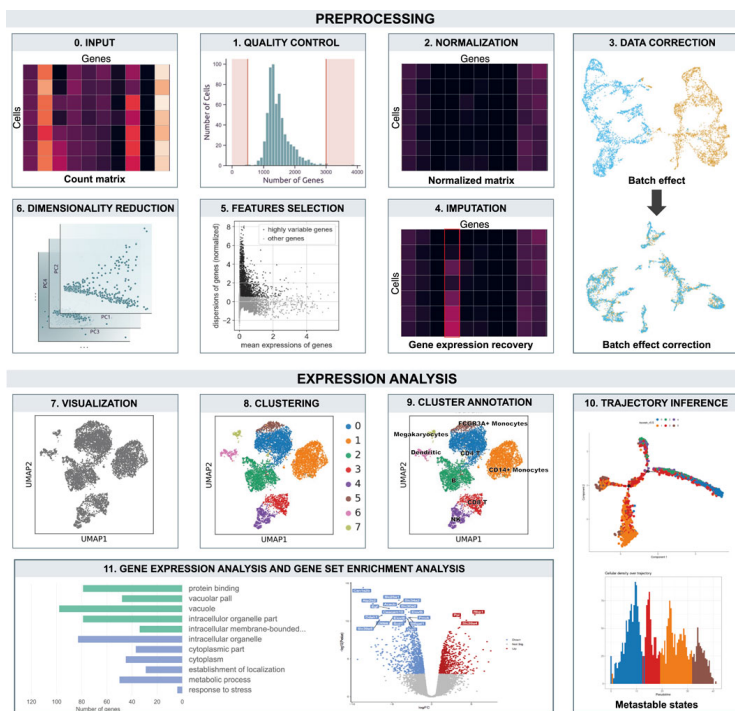
Λαμβάνοντας υπόψη όλα τα παραπάνω, θα παρουσιαστεί μια επισκόπηση της υπολογιστικής ροής εργασίας, αναφέροντας τα διαθέσιμα εργαλεία για την επίτευξη του κάθε βήματος και επισημαίνοντας τις βασικές τεχνολογικές εξελίξεις που επέτρεψαν την ανάλυση αυτού του συνεχώς αυξανόμενου όγκου δεδομένων, καθιστώντας την τεχνολογία scRNA-seq ένα εύχρηστο εργαλείο στη βιοϊατρική έρευνα.

2.2. Γενικές πληροφορίες και ροή εργασίας

Η ταχεία τεχνολογική εξέλιξη που επέτρεψε την παράλληλη ανάλυση χιλιάδων κυττάρων, προωθώντας την εξάπλωση των τεχνικών scRNA-seq, συνοδεύτηκε από την ανάπτυξη νέων αγωγών ανάλυσης δεδομένων ικανών να διαχειριστούν τόσο μεγάλο όγκο δεδομένων. Η μαθηματική αναπαράσταση αυτών των μαζικών συνόλων δεδομένων είναι ένα μητρώο έκφρασης (expression matrix), που ορίζεται από τον αριθμό των ανιχνευμένων γονιδίων και των παρατηρούμενων κυττάρων αντίστοιχα. Η διαδικασία ξεκινά με τον έλεγχο ποιότητας ανάγνωσης. Τα αρχεία FastQ [2] που εξάγονται από το sequencer αξιολογούνται μέσω εργαλείων ελέγχου ποιότητας, όπως το FastQC. Προσαρμοσμένοι αγωγοί όπως ο Cell Ranger,

UMI-tools, scPipe και zUMIs, αναπτύχθηκαν για την εκτέλεση αυτών των προκαταρκτικών βημάτων. Εναλλακτικά, οι ερευνητές μπορούν να δημιουργήσουν τις δικές τους ροές εργασίας συνδυάζοντας μεμονωμένες μεθόδους που αντιμετωπίζουν κάθε μία από τις προαναφερθείσες εργασίες.

Επιπλέον, εάν οι αναγνώσεις είναι UMI (Unique Molecular Identifier) -tagged [3], διατηρούνται μόνο οι γραμμωτοί κώδικες κυττάρων (cell barcodes) που αντιπροσωπεύουν άθικτα μεμονωμένα κύτταρα. Η πιο ξεκάθαρη προσέγγιση για την αξιολόγηση είναι ο υπολογισμός ενός ορίου συγκεκριμένου συνόλου δεδομένων του ελάχιστου αριθμού UMI που απαιτείται για να θεωρηθεί ένας γραμμικός κώδικας ως κύτταρο. Τα προκύπτοντα κύτταρα εξακολουθούν να εμφανίζουν ανεπιθύμητες προκαταλήψεις. Όλες οι διαδικασίες που εμπλέκονται στην αφαίρεση της μεροληψίας ορίζουν την προεπεξεργασία, η οποία συνίσταται στον ποιοτικό έλεγχο, την ομαλοποίηση, τη διόρθωση παρτίδων, την επιλογή χαρακτηριστικών και τη μείωση των διαστάσεων. Όλα αυτά τα βήματα είναι προπαρασκευαστικά για την ακόλουθη ανάλυση έκφρασης, που χρησιμοποιείται για την εξαγωγή βιολογικών γνώσεων και την περιγραφή του υποκείμενου βιολογικού συστήματος.



Εικόνα 1: Βήματα scRNA-seq

2.3. Ποιοτικός έλεγχος

Πριν πραγματοποιηθεί ανάλυση του μητρώου έκφρασης, πρέπει να πραγματοποιηθεί αξιολόγηση της μοναδικότητας κάθε γραμμωτού κώδικα και της βιωσιμότητας των κυττάρων. Για το σκοπό αυτό, είναι σημαντικό να λαμβάνεται υπόψη ότι ορισμένα δείγματα μπορεί να περιέχουν περισσότερα από 1 κύτταρο ή καθόλου κύτταρα, καθιστώντας το διπλό, πολλαπλό ή άδειο δείγμα. Επιπλέον, τα κύτταρα μπορεί να πεθαίνουν ή να καταστρέφονται κατά τη διάρκεια της απομόνωσης, παραποιώντας τη σύνθεση του δείγματος. Έτσι, θα πρέπει να πραγματοποιείται διαδικασία φιλτραρίσματος.

Μια πιθανή λύση είναι να προσδιοριστούν αυτά τα κύτταρα αξιολογώντας τρεις πτυχές των δεδομένων: τον αριθμό των μετρήσεων ανά κύτταρο/γραμμωτό κώδικα (βάθος μέτρησης), τον αριθμό των γονιδίων ανά κύτταρο/γραμμωτό κώδικα και το κλάσμα των μετρήσεων από μιτοχονδριακά γονίδια ανά κύτταρο/γραμμωτό κώδικα. Τα όρια είναι αυθαίρετα με βάση τα γενικά χαρακτηριστικά των ίδιων των δεδομένων.

Ωστόσο, μια παρερμηνεία αυτών των δεδομένων θα μπορούσε να οδηγήσει σε λάθος φιλτράρισμα, καθώς σε ορισμένες περιπτώσεις μια απόκλιση σε μία από αυτές τις τιμές μπορεί να σχετίζεται με μια συγκεκριμένη κυτταρική κατάσταση, όπως βαριά αναπνοή (υψηλός αριθμός μιτοχονδρίων), ηρεμία (χαμηλός αριθμός, λίγα γονίδια) και μεγαλύτερο μέγεθος (υψηλός αριθμός).

Ο ποιοτικός έλεγχος μπορεί επίσης να περιλαμβάνει ένα βήμα φιλτραρίσματος γονιδίων, δεδομένου ότι τα γονίδια που εκφράζονται σε λίγα κύτταρα δεν είναι ενημερωτικά για την κυτταρική ετερογένεια. Το κατώτατο όριο είναι και πάλι αυθαίρετο, αλλά κατ' αρχήν θα πρέπει να κλιμακώνεται με τον αριθμό των κυττάρων στο σύνολο δεδομένων.

2.4. Κανονικοποίηση

Μέσω του ποιοτικού ελέγχου αφαιρούνται πηγές ανεπιθύμητων και ανακριβών πληροφοριών. Ωστόσο, το σύνολο δεδομένων εξακολουθεί να επηρεάζεται από πολλαπλές προκαταλήψεις λόγω τεχνικής και βιολογικής μεταβλητότητας. Οι πηγές που είναι υπεύθυνες για τέτοια γεγονότα θα μπορούσαν να είναι, για παράδειγμα, η αποτελεσματικότητα της καταγραφής, η ενίσχυση και η ελλιπής αλληλούχιση. Η συνέπεια είναι μια μεταβολή των αριθμών που καθιστούν τα κύτταρα ασύγκριτα .

Η κανονικοποίηση αντιμετωπίζει αυτό το ζήτημα π.χ. με την κλιμάκωση των δεδομένων καταμέτρησης για να ληφθούν σωστές σχετικές αφθονίες γονιδιακής έκφρασης μεταξύ των κυττάρων. Οι διαθέσιμες μέθοδοι μπορεί να είναι γραμμικές ή μη γραμμικές: μια γραμμική προσέγγιση περιλαμβάνει την εκτίμηση των παραγόντων μεγέθους με βάση μια γραμμική παλινδρόμηση έναντι των γονιδίων, ενώ οι μη γραμμικές μέθοδοι συνήθως εφαρμόζουν παραμετρική μοντελοποίηση σε δεδομένα καταμέτρησης και συσχετίζουν τεχνικές και βιολογικές πηγές μεταβλητότητας για τη διόρθωση και των δύο .

Η πιο κοινή προσέγγιση ομαλοποίησης είναι η κλιμάκωση βάθους καταμέτρησης με "αριθμούς ανά εκατομμύριο" (CPM) [4], η οποία λειτουργεί διαιρώντας τους αριθμούς γονιδίων με τον συνολικό αριθμό χαρτογραφημένων αναγνώσεων ανά δείγμα και πολλαπλασιάζοντας με 1×10^6 . Το CPM εμπίπτει στις μεθόδους κανονικοποίησης γραμμικής καθολικής κλιμάκωσης και υποθέτει ότι όλα τα κύτταρα στο σύνολο δεδομένων περιείχαν αρχικά ίσο αριθμό μορίων mRNA και οι διαφορές βάθους καταμέτρησης προκύπτουν μόνο λόγω δειγματοληψίας. Οι παραλλαγές αυτής της μεθόδου κλιμακώνουν τους παράγοντες μεγέθους με διαφορετικούς παράγοντες.

Ένας άλλος κρίσιμος παράγοντας για την ομαλοποίηση είναι η παρουσία spike – ins ή UMIs. Με την προσθήκη γνωστών συγκεντρώσεων εξωτερικών μεταγραφών, που ονομάζονται ακίδες, είναι δυνατόν να αξιολογηθεί η παρουσία τεχνικών τεχνουργημάτων, αναζητώντας διαφορές μεταξύ της παρατηρούμενης και της αναμενόμενης έκφρασής τους. Με τον υπολογισμό ενός κυτταρικού ειδικού παράγοντα και με την εφαρμογή αυτού του παράγοντα σε ενδογενή γονίδια, μπορούν να ληφθούν κανονικοποιημένες εκτιμήσεις έκφρασης. Σε αντίθεση με τα spike-ins, τα UMIs είναι ευκολότερα στο χειρισμό, δεδομένου ότι συνδέονται με μεμονωμένες μεταγραφές πριν από την PCR [5], καθιστώντας κάθε μόριο μοναδικό.

Επίσης, τα γονίδια μπορούν να ομαλοποιηθούν ώστε να είναι συγκρίσιμα μεταξύ των κυττάρων. Οι αριθμοί γονιδίων μπορούν να κλιμακωθούν ώστε να έχουν μηδενικό μέσο όρο και διακύμανση, καθιστώντας τα γονίδια εξίσου σταθμισμένα.

2.5. Διόρθωση Παρτίδας

Για την απαλλαγή από τις υφιστάμενες προκαταλήψεις στα δεδομένα, είναι δυνατόν να πραγματοποιηθεί διόρθωση δεδομένων και παρτίδων. Επί του παρόντος, διάφορα εργαλεία μπορούν να ολοκληρώσουν αυτές τις εργασίες με διαφορετικές προσεγγίσεις. Για παράδειγμα, σε μελέτες προσανατολισμένες στην ανάπτυξη, η υποχώρηση του αποτελέσματος του κυτταρικού κύκλου θα μπορούσε να αποκαλύψει τα επιθυμητά βιολογικά. Για το σκοπό αυτό, μέθοδοι όπως το Scanpy και το Seurat εφαρμόζουν λειτουργίες για να βαθμολογήσουν τις φάσεις του κυτταρικού κύκλου και να υποχωρήσουν γραμμικά τη βιολογική τους επίδραση. Εναλλακτικά, προσαρμοσμένα εργαλεία βασισμένα σε σύνθετα μοντέλα, όπως το f-scLVM, είναι διαθέσιμα. Μερικές φορές, επίσης, η μεροληψία που παράγεται από τις διαφορές στο μέγεθος των κυττάρων, αν δεν διορθωθεί αρκετά μέσω της ομαλοποίησης, θα μπορούσε να μετριάσει περαιτέρω για να τονίσει τα σήματα που σχετίζονται με την ανάπτυξη.

Η διόρθωση όσο αφορά τις βιολογικές προκαταλήψεις, ωστόσο, δεν είναι πάντα απαραίτητη ή χρήσιμη, καθώς μπορούν να αποφευχθούν μέσω συλλογισμένου πειραματικού σχεδιασμού ή επειδή μπορούν να σχετίζονται με τη βιολογική διαδικασία ενδιαφέροντος. Η ίδια παρατήρηση ισχύει εν μέρει και για εκείνες τεχνικής φύσης. Στην πραγματικότητα, ακόμη και σε αυτή την περίπτωση ένας έξυπνος πειραματικός σχεδιασμός επιτρέπει τη μείωση της επιρροής τους, αλλά, εάν υπάρχουν, δεν έχουν καμία συσχέτιση με τα βιολογικά σήματα, οπότε πρέπει να μετριάστούν. Αυτή η διαδικασία, που ονομάζεται διόρθωση παρτίδας, μπορεί να διεξαχθεί μεταξύ δειγμάτων και κυττάρων του ίδιου πειράματος μέσω γραμμικών μοντέλων ή μεταξύ διαφορετικών συνόλων δεδομένων που προέρχονται από πολλαπλές πειραματικές ρυθμίσεις μέσω μη γραμμικών μοντέλων.

Μία από τις πιο κοινές γραμμικές μεθόδους είναι η ComBat [6] που λαμβάνει υπόψη την επίδραση της παρτίδας στη μέση τιμή και τη διακύμανση του συνόλου δεδομένων, με πολύ καλή απόδοση.

Εάν οι διαφορές στα σύνολα δεδομένων είναι πιο έντονες, τα γραμμικά μοντέλα θα μπορούσαν να συγχέουν τις ενδοτεχνικές και διατεχνικές και βιολογικές προκαταλήψεις και υπό αυτές τις συνθήκες μη γραμμικά μοντέλα που εφαρμόζονται σε εργαλεία, έχει αποδειχθεί ότι αντιμετωπίζουν το ίδιο ζήτημα και εξομαλύνουν τις ανεπιθύμητες και παραπλανητικές διαφορές.

2.6. Καταλογισμός (Imputation)

Οι πληροφορίες που αποθηκεύονται σε ένα σύνολο δεδομένων ενός κυττάρου έχουν πολύ αραιή φύση. Με μαθηματικούς όρους, μεταφράζεται σε ένα μητρώο γεμάτο με μηδενικά. Πολλές προσεγγίσεις ομαλοποίησης δεν τα αφαιρούν, υποθέτοντας ότι αντιπροσωπεύουν ελλείπουσες τιμές για να ληφθούν υπόψη στους υπολογισμούς. Ωστόσο, η μείωση του αριθμού τους θα μπορούσε να μειώσει τον θόρυβο, βελτιώνοντας την εκτίμηση των συσχετίσεων γονιδίων-γονιδίων.

Επί του παρόντος, πολλά εργαλεία είναι διαθέσιμα για την επίτευξη αυτού του έργου και τα καλύτερα αποδοτικά βασίζονται κυρίως σε αλγόριθμους βαθιάς μάθησης. Σε αυτή την κατηγορία εμπίπτουν οι μέθοδοι DeepImpute και Deep Count Autoencoder network (DCA) [7]. Η πρώτη μέθοδος χρησιμοποιεί πολύ συσχετισμένα γονίδια των γονιδίων-στόχων για να καταλογίσει τις ελλείπουσες τιμές, ενώ η δεύτερη μέθοδος μπορεί να συλλάβει τη μη γραμμική συσχέτιση γονιδίου-γονιδίου. Η εφαρμογή τους αποδείχθηκε ότι βελτιώνει την απόδοση στην ομαδοποίηση κυττάρων.

Ωστόσο, κατά την εφαρμογή της ανάκτησης έκφρασης, πρέπει να ληφθεί υπόψη ότι καμία μέθοδος δεν είναι τέλεια. Έτσι, οποιαδήποτε μέθοδος μπορεί να εισάγει θόρυβο στα δεδομένα. Πράγματι, έχουν αναφερθεί ψευδή σήματα συσχέτισης ως αποτέλεσμα της ανάκτησης έκφρασης.

Υπό το πρίσμα αυτό, είναι δύσκολο να εκτιμηθεί εάν ο καταλογισμός θα επιτύχει σε μια συγκεκριμένη εφαρμογή. Μια λογική προσέγγιση θα ήταν να καταλογιστεί η οπτικοποίηση και

να αποφευχθεί η δημιουργία υποθέσεων κατά τη διάρκεια της διερευνητικής ανάλυσης δεδομένων.

2.7. Επιλογή χαρακτηριστικών και μείωση διαστάσεων

Αφού έχει ολοκληρωθεί το βήμα του καθαρισμού των δεδομένων, ένα σύνολο δεδομένων scRNA-seq μπορεί ακόμα να περιέχει έως και 15.000 γονίδια. Ένα τόσο μεγάλο και πολυδιάστατο σύνολο είναι, ωστόσο, δύσκολο να διαχειριστεί και να απεικονιστεί. Για τους λόγους αυτούς, υπόκειται σε μείωση διαστάσεων.

Για να περάσουμε από αυτή τη διαδικασία, είναι σημαντικό να έχουμε κατά νου ότι πολλά υπολειμματικά γονίδια δεν αντιπροσωπεύουν τη μεταβλητότητα των δεδομένων, η οποία αποτελεί βασικό χαρακτηριστικό για τη διερεύνηση της ετερογένειας του δείγματος και έτσι μπορούμε να τα θεωρήσουμε μη ενημερωτικά και αγνοούμενα. Αυτή η διαδικασία ονομάζεται επιλογή χαρακτηριστικών. Ένας κοινός τρόπος για να επιτευχθεί αυτό το αποτέλεσμα είναι να αναζητηθούν πολύ μεταβλητά γονίδια (HVGs) [8] με την αντικατάστασή τους με τη μέση έκφρασή τους και τη διατήρηση εκείνων με την υψηλότερη αναλογία μέσης τιμής προς διακύμανση σε κάθε bin. Μέθοδοι όπως το Scanpy και το Cell Ranger [9] εφαρμόζουν λειτουργίες για τον ορισμό των HVGs ξεκινώντας από δεδομένα μετασχηματισμένα με λογάριθμο, ενώ άλλες όπως το Seurat εργάζονται στους ακατέργαστους αριθμούς. Συνήθως, επιλέγονται μεταξύ 1.000 και 5.000 HVGs.

Η ταυτοποίησή τους είναι επίσης ζωτικής σημασίας για να προχωρήσουμε στην ακόλουθη μείωση των διαστάσεων. Πράγματι, κοινές μέθοδοι όπως η ανάλυση κύριων συστατικών (PCA) [10] επωφελούνται από τη χρήση HVGs για τον ορισμό των μειωμένων συστατικών που χρησιμοποιούνται για τη σύνοψη των χαρακτηριστικών του συνόλου δεδομένων σε ένα χώρο χαμηλής διάστασης. Αυτό είναι δυνατό μέσω μιας γραμμικής προσέγγισης που μετατρέπει ένα σύνολο συσχετισμένων μεταβλητών σε μικρότερο αριθμό μη συσχετισμένων μεταβλητών, που ονομάζονται κύρια συστατικά, διατηρώντας όσο το δυνατόν μεγαλύτερη διακύμανση των δεδομένων.

Η PCA είναι μια τεχνική που προέρχεται από το πεδίο της γραμμικής άλγεβρας και μπορεί να χρησιμοποιηθεί ως τεχνική προετοιμασίας δεδομένων για να δημιουργήσει μια προβολή ενός συνόλου δεδομένων πριν από την εκπαίδευση ενός μοντέλου. Πράγματι, για σύνθετα σύνολα δεδομένων των οποίων η δομή δεν μπορούσε να συλληφθεί, μη γραμμικές μέθοδοι συνδυασμού όπως η T-distributed Stochastic Neighbour Embedding (t-SNE) [11] και η ομοιόμορφη προσέγγιση και προβολή (UMAP) [12] αποδίδουν καλύτερα, εκμεταλλευόμενοι τα δεδομένα PCA.

2.8. Οπτικοποίηση και ομαδοποίηση

Οι μη γραμμικές μέθοδοι χρησιμοποιούνται συνήθως για τη δημιουργία ενός διδιάστατου plot που συνοψίζει ένα σύνολο δεδομένων scRNA-seq από μεγαλύτερο αριθμό σημαντικών συστατικών. Οι μέθοδοι t-SNE και το UMAP είναι δύο τυπικές λύσεις για την επίτευξη αυτού του έργου και εφαρμόζονται σχεδόν σε όλες τις εργαλειοθήκες επεξεργασίας δεδομένων scRNA-seq. Η T-SNE παίρνει ένα σύνολο δεδομένων υψηλής διάστασης και το μειώνει σε ένα γράφημα χαμηλής διάστασης που επικεντρώνεται στην καταγραφή της τοπικής ομοιότητας σε βάρος της καθολικής δομής. Η UMAP, αντίθετα, τείνει να ευνοεί πλήρως συνδεδεμένες αναπαραστάσεις του συνόλου δεδομένων χρησιμοποιώντας ένα δίκτυο πλησιέστερου γείτονα κυττάρου-κυττάρου για να εκτιμήσει στη συνέχεια μια χαμηλής διάστασης ενσωμάτωση των δεδομένων. Η τελευταία αντικαθιστά σε μεγάλο βαθμό την πρώτη, αν και διαφορετικές αναπαραστάσεις θα μπορούσαν να δώσουν διαφορετικές ιδέες. Σε αυτή την προοπτική, είναι καλό να γνωρίζουμε ότι υπάρχουν επίσης χάρτες διάχυσης και μέθοδοι που βασίζονται σε διαμερίσματα για την απεικόνιση σύνθετων δεδομένων με διαφορετικούς τρόπους και για διαφορετικές εφαρμογές, π.χ. οι χάρτες διάχυσης είναι αποτελεσματικοί για να εξάγουν συμπεράσματα σε αναλύσεις τροχιάς, ενώ οι μέθοδοι που βασίζονται σε διαμερίσματα προσεγγίζουν την τοπολογία των δεδομένων χρησιμοποιώντας συστάδες για να παράγουν μια απλοποιημένη απεικόνιση των δεδομένων, χρήσιμη με πολύ μεγάλα σύνολα δεδομένων.

Η ομαδοποίηση πραγματοποιείται συνήθως μέσω των αλγορίθμων Louvain και Leiden. Ο στόχος αυτού του βήματος είναι να ορίσει ομάδες κυττάρων με παρόμοια προφίλ έκφρασης,

επειδή αυτές οι ομάδες θα μπορούσαν να αντιπροσωπεύουν τύπους κυττάρων, ενδιάμεσες κυτταρικές καταστάσεις ή άλλες ενδιαφέρουσες πτυχές των δεδομένων.

Και οι δύο μέθοδοι βασίζονται στην προσέγγιση K-Nearest Neighbor (KNN graph) όπου τα κύτταρα αντιπροσωπεύονται ως κόμβοι σε ένα γράφημα, το καθένα συνδεδεμένο με τα K πιο παρόμοια κύτταρα του, που λαμβάνονται χρησιμοποιώντας ευκλείδειες αποστάσεις στον χώρο έκφρασης, έτσι ώστε οι περιοχές έκφρασης με πυκνά δείγματα να αντιπροσωπεύονται ως πυκνά συνδεδεμένες περιοχές.

Η ομαδοποίηση μπορεί επίσης να πραγματοποιηθεί σε πολλαπλές αναλύσεις για την επιθεώρηση δεδομένων σε διαφορετικά επίπεδα λεπτομέρειας (δηλαδή, περισσότερες συστάδες μικρότερων διαστάσεων). Επιπλέον, οι ομάδες που προκύπτουν μπορούν να υπο - ομαδοποιηθούν επαναληπτικά για να επιτρέψουν την ταυτοποίηση των κυτταρικών καταστάσεων που συλλαμβάνονται μέσα στο ίδιο σύμπλεγμα.

2.9. Σχολιασμός συστάδας (Cluster Annotation)

Μόλις καθοριστούν οι συστάδες, είναι καιρός να προσδιοριστούν οι αντιπροσωπευόμενοι κυτταρικοί πληθυσμοί. Αυτό μπορεί να γίνει με τον ορισμό των γονιδιακών υπογραφών τους μέσω της ταυτοποίησης γονιδίων - δεικτών. Για το σκοπό αυτό, οι δοκιμές DE εφαρμόζονται συνήθως μεταξύ δύο ομάδων που αντιπροσωπεύουν το σύμπλεγμα και το υπόλοιπο σύνολο δεδομένων. Στη συνέχεια, απλές στατιστικές δοκιμές όπως η δοκιμή Wilcoxon rank-sum ή η t-test χρησιμοποιούνται για την κατάταξη των παραγόμενων γονιδίων με βάση τη διαφορά στην έκφρασή τους. Τα κορυφαία γονίδια από την αντίστοιχη στατιστική δοκιμής θεωρούνται ως γονίδια - δείκτες.

Οι συστάδες μπορούν επίσης να σχολιαστούν συγκρίνοντας γονίδια - δείκτες από το σύνολο δεδομένων με εκείνα από σύνολα δεδομένων αναφοράς μέσω δοκιμών εμπλουτισμού, του δείκτη Jaccard ή άλλων στατιστικών αλληλεπικάλυψης. Επίσης, αυτοματοποιημένες μέθοδοι

όπως η single cell NET είναι διαθέσιμες για να ολοκληρώσουν αυτό το βήμα και να επιταχύνουν τη διαδικασία σχολιασμού, αν και προτείνεται πάντα μια χειροκίνητη αναθεώρηση λόγω της πλαστικότητας των κυτταρικών καταστάσεων που μερικές φορές θα μπορούσαν να συγχέονται με άλλες.

2.10. Ανάλυση τροχιάς

Η ομαδοποίηση κυττάρων μερικές φορές δεν είναι η κατάλληλη στρατηγική για τη μελέτη ενός συνόλου δεδομένων. Πολλές βιολογικές διεργασίες, που χαρακτηρίζουν ένα σύνολο δεδομένων, δεν μπορούν να περιγραφούν μέσω διακριτής ταξινόμησης αλλά μάλλον με πιο συνεχή τρόπο. Για να επιτευχθεί αυτό το αποτέλεσμα πρέπει να εφαρμόσουμε γονιδιακά δυναμικά μοντέλα ικανά να ταξινομήσουν τα κύτταρα κατά μήκος ενός άξονα που ορίζει τη χρονική διαδικασία, γνωστή και ως ψευδοχρόνος. Αυτός ο τύπος προσέγγισης χρησιμοποιείται συνήθως για τη μελέτη διαδικασιών όπως η ανάπτυξη και η διαφοροποίηση και ονομάζεται ανάλυση τροχιάς.

Αρκετές μέθοδοι είναι επί του παρόντος διαθέσιμες για να συμπεράνουν τροχιές αυξανόμενης πολυπλοκότητας, από απλές γραμμικές ή διακλαδικές διαδρομές σε σύνθετα γραφήματα, δέντρα ή πιο περίπλοκες τροχιές.

Συνήθως, αυτοί οι αλγόριθμοι λαμβάνουν τα μειωμένα ή διορθωμένα δεδομένα ως είσοδο για να ελαχιστοποιήσουν την τεχνική παραλλαγή και να συλλάβουν μόνο τη βιολογική, εκμεταλλευόμενοι επίσης τα HVGs, τα οποία χρησιμοποιούνται για τον καθορισμό των διαδοχικών καταστάσεων που προέρχονται από μεταγραφικές αποστάσεις από ένα κύτταρο. Καμία από τις διαθέσιμες μεθόδους δεν έχει αποδειχθεί ότι υπεραποδίδει έναντι των άλλων για όλα τα είδη τροχιών, αν και διαφορετικές προσεγγίσεις ωφελούν διαφορετικούς σκοπούς, όπως φαίνεται σε προηγούμενες συγκριτικές μελέτες .

Για παράδειγμα, το εργαλείο Slingshot [13] αποδείχθηκε ότι είχε καλύτερες επιδόσεις όταν συνήγαγε γραμμικές ή πολλαπλές τροχιές, σε αντίθεση με το Monocle2 [14], το οποίο δίνει τα καλύτερα αποτελέσματα σε πιο σύνθετες και διακλαδισμένες καταστάσεις.

Η προαναφερθείσα εργαλειοθήκη rpyhton προσφέρει επίσης την ευκαιρία να συμβιβάσει τις πληροφορίες που προέρχονται από την ομαδοποίηση και την εξαγωγή τροχιάς, μέσω του

αλγορίθμου αφαίρεσης γραφήματος με βάση το partition (PAGA). Πιο συγκεκριμένα, χρησιμοποιώντας ένα στατιστικό μοντέλο για αλληλεπιδράσεις συστάδων κυττάρων, το PAGA [15] τοποθετεί μια ακμή μεταξύ κόμβων συστάδων των οποίων τα κύτταρα είναι πιο παρόμοια από το αναμενόμενο, δημιουργώντας έναν χάρτη που αντιπροσωπεύει τη στατική και δυναμική φύση των δεδομένων.

Καθώς η ανάλυση τροχιάς ασχολείται με τον τρόπο με τον οποίο τα κύτταρα στο δείγμα μας αλλάζουν σύμφωνα με τον ψευδοχρόνο, γίνεται δυνατόν να οριστούν οι μεταγραφικές καταστάσεις της διαδικασίας αξιολόγησης της πυκνότητας της περιοχής. Οι πυκνές περιοχές των κυττάρων αντιπροσωπεύουν τις λεγόμενες «μεταστατικές καταστάσεις», που μπορούν να απεικονιστούν μέσω ιστογραμμάτων.

Δυστυχώς, λίγες από τις προαναφερθείσες μεθόδους περιλαμβάνουν αξιολόγηση της αβεβαιότητας στο μοντέλο τους, οπότε τα προβλεπόμενα αποτελέσματα θα πρέπει να επιβεβαιωθούν με εναλλακτικές προσεγγίσεις για την αποφυγή μεροληψίας. Ένας κοινός τρόπος για να επιτευχθεί αυτός ο στόχος είναι να συναχθεί η δυναμική του χρόνου. Επί του παρόντος, υπάρχουν δύο προσεγγίσεις μοντελοποίησης, το αρχικά προτεινόμενο μοντέλο «σταθερής κατάστασης» και το μεταγενέστερο εκτεταμένο δυναμικό μοντέλο.

2.11. Ανάλυση γονιδιακής έκφρασης

Μόλις αξιολογηθεί η φύση κάθε συμπλέγματος, η εστίαση στην έκφραση γονιδίων μπορεί να μας δώσει μια πολύ ευρύτερη ιδέα για διαδικασίες και μηχανισμούς που διαφέρουν μεταξύ τους. Σε αυτή την προοπτική, εργαλεία όπως η ανάλυση DE και η ανάλυση εμπλουτισμού γονιδιακών συνόλων (GSEA) [16] μπορούν να μας βοηθήσουν να διερευνήσουμε τη μοριακή μεταβλητότητα που προέρχεται από διαφορετικές πειραματικές (ιατρική θεραπεία) ή βιολογικές (διαφορετικές κυτταρικές σειρές) συνθήκες.

Οι μέθοδοι DE προέρχονται από ανάλυση δεδομένων μαζικής αλληλουχίας, όπου μερικά δείγματα συγκρίνονται για να κατανοήσουν τις μοριακές συνέπειες διαφορετικών πειραματικών συνθηκών. Σε μονοκύτταρες ρυθμίσεις, οι μεταβλητές που εξετάζονται αυξάνονται καθώς

αυξάνεται ο αριθμός των κυττάρων που εξετάζονται, λόγω της μεταβλητότητας από κύτταρο σε κύτταρο και των προκαταλήψεων.

Προσαρμοσμένα εργαλεία όπως το MAST ή scCODE [17] είναι διαθέσιμα για να χειριστούν αυτά τα χαρακτηριστικά και να εκτελέσουν DE σε μεγάλα σύνολα δεδομένων ενός κυττάρου σε εύλογους χρόνους, ωστόσο, μαζικά DE εργαλεία, όπως το DESeq2 και το EdgeR [18], έχουν αποδειχθεί ότι ξεπερνούν ορισμένα αντίστοιχα μονοκύτταρα εάν βαθμονομηθούν σωστά, αλλά διαρκούν πολύ. Τα μη διορθωμένα δεδομένα προτιμούνται για αυτές τις εφαρμογές, οπότε είναι ζωτικής σημασίας να ληφθούν υπόψη οι παράγοντες σύγχυσης για να εκτελεστεί μια ισχυρή εκτίμηση των διαφορικά εκφρασμένων γονιδίων.

Το αποτέλεσμα της δοκιμής συνίσταται σε έναν μακρύ κατάλογο γονιδίων που εκφράζονται διαφορικά μεταξύ δύο ή περισσότερων συνθηκών, μερικές φορές δύσκολο να ερμηνευθούν με ουσιαστικό τρόπο. Για να ξεπεραστεί αυτός ο περιορισμός, μπορούν να αναλυθούν τα ομαδοποιώντας σε σύνολα με βάση κοινά χαρακτηριστικά. Αυτή η προσέγγιση, που ονομάζεται GSEA, ελέγχει εάν αυτά τα χαρακτηριστικά υπερεκπροσωπούνται στον κατάλογο των υποψήφιων γονιδίων και βασίζεται στη χρήση βάσεων δεδομένων όπως η οντολογία γονιδίων, η κοινοπραξία οντολογίας γονιδίων, 2017), KEGG, String, και Reactome. Εργαλεία όπως το gseapy και το biomaRt είναι διαθέσιμα για την εκτέλεση αυτής της εργασίας μέσω πολλαπλών δοκιμών, αναζητώντας τις αναφερόμενες βάσεις δεδομένων.

2.12. Ζητήματα πειραματικού σχεδιασμού

Η scRNA-seq έχει ανοίξει νέους δρόμους για τον χαρακτηρισμό της ετερογένειας σε μια μεγάλη ποικιλία κυτταρικών συστημάτων, επιτρέποντας τη λήψη δεδομένων σε ολόκληρο το μεταγραφικό από μεμονωμένα κύτταρα. Αν και το προφίλ γονιδιακής έκφρασης σε επίπεδο ενός κυττάρου αποκάλυψε μια άνευ προηγουμένου ποικιλία κυτταρικών τύπων και υποπληθυσμών που ήταν αόρατοι με τις παραδοσιακές πειραματικές τεχνικές, εισήγαγε νέες προκλήσεις λόγω της εγγενούς φύσης των δεδομένων.

Πράγματι, τα σύνολα δεδομένων scRNA-seq παρουσιάζουν αυξημένη μεταβλητότητα, σύνθετες κατανομές έκφρασης και αφθονία μηδενικών σε σύγκριση με εκείνα που παράγονται σε

πειράματα "bulk", καθιστώντας δύσκολη τη δημιουργία ευρέως εφαρμόσιμων πειραματικών σχεδίων. Υπό το πρίσμα αυτό, κάθε πείραμα απαιτεί από τον χρήστη να λαμβάνει τεκμηριωμένες αποφάσεις πριν προχωρήσει σε ένα σκεπτόμενο σχέδιο, το οποίο πρέπει να ικανοποιεί τρεις αρχές: αναπαραγωγή, τυχαιοποίηση και δημιουργία μπλοκ.

Για να προετοιμαστεί ένα πείραμα, τηρώντας τέτοιες αρχές, είναι καλό να δημιουργηθεί ένας ισορροπημένο σχεδιασμός μπλοκ, στον οποίο τα δείγματα που συλλέγονται από πολλαπλές συνθήκες κατανέμονται ομοιόμορφα, προκειμένου να μειωθεί η τεχνική παραλλαγή και να μην συγχέεται με τη βιολογική. Ο ισορροπημένος σχεδιασμός επιτρέπει την παράκαμψη του βήματος διόρθωσης παρτίδας στην υπολογιστική ανάλυση, μειώνοντας τους υπολογιστικούς χρόνους και την παρέμβαση του χρήστη στα δεδομένα.

Οι πειραματικές εκτιμήσεις σχεδιασμού θα επηρεαστούν επίσης από τα διάφορα πρωτόκολλα και πλατφόρμες που διατίθενται για το scRNA-seq. Για παράδειγμα, οι μέθοδοι πλήρους μήκους προσφέρουν διαφορετικό τρόπο εξερεύνησης των χαρακτηριστικών του δείγματος.

Επιπλέον, σε περίπτωση που δεν υπάρχει προηγούμενη γνώση σχετικά με την ετερογένεια του κυτταρικού πληθυσμού, μια πρακτική λύση είναι να εκτελεστεί η μελέτη με υψηλό αριθμό κυττάρων και χαμηλότερο βάθος αλληλουχίας και στη συνέχεια να πραγματοποιηθεί προκαθαρισμός των κυττάρων.

2.13. Εφαρμογές

Η σύγχρονη γνώση της κυτταρικής και της Μοριακής Βιολογίας προέρχεται σε μεγάλο βαθμό από πειράματα RNA-seq που επέτρεψαν την κατανόηση της πολυπλοκότητας της δυναμικής που είναι υπεύθυνη για μεταβολικές μεταβολές, τροφοδοτώντας πολλές ανακαλύψεις και καινοτομίες στον τομέα της ιατρικής τα τελευταία χρόνια.

Η εξέλιξη τέτοιων τεχνικών καθοδηγήθηκε από την ανάπτυξη πρωτοκόλλων και συσκευών ικανών να εξάγουν μεταγραφικές πληροφορίες από έναν συνεχώς αυξανόμενο αριθμό μεμονωμένων κυττάρων, θέτοντας τις βάσεις για μελέτες προσανατολισμένες στην ετερογένεια.

Η ευκαιρία να τεμαχιστεί ένα δείγμα άνοιξε νέες προοπτικές σε κλινικές μελέτες προσανατολισμένες στην ανακάλυψη σπάνιων κυτταρικών πληθυσμών που εμπλέκονται στην εμφάνιση και εξέλιξη ασθενειών όπως οι όγκοι. Μια απόδειξη αυτού του ισχυρισμού από μια μελέτη , η οποία εξέτασε την αρχιτεκτονική σύνθεσης των δειγμάτων μελανώματος και γλοιοβλαστώματος σε επίπεδο ενός κυττάρου. Μετά από την μελέτη αυτή, ένας αυξανόμενος αριθμός μελετών και ερευνητών έχουν αρχίσει να εκμεταλλεύονται την τεχνική για να χαρακτηρίσουν επιτυχώς τους κυτταρικούς πληθυσμούς σε μια ποικιλία όγκων, καθορίζοντας το ρόλο τους στη διαδικασία της νόσου και την ταυτότητά τους μέσω της ανάθεσης γονιδιακών υπογραφών. Άλλες συνεισφορές στον τομέα προέρχονται από την ενσωμάτωση της ανίχνευσης scRNA-seq και Copy Number Variant (CNV) [19]. Μια μελέτη εφάρμοσε με επιτυχία αυτήν την τεχνική για να αποκτήσει νέες γνώσεις σχετικά με την ενδο - και διαπροσωπική, χωρική, λειτουργική και γονιδιωματική ετερογένεια στα κύτταρα μελανώματος, καθώς και λεπτομέρειες που σχετίζονται με το μικροπεριβάλλον του όγκου και τα κύτταρα που το κατοικούν, επικυρώνοντας την παρουσία ενός αδρανούς πληθυσμού ανθεκτικού στα φάρμακα .

Ομοίως, το 2018, μια επιπλέον μελέτη εκμεταλλεύτηκε τα CNVs και την απώλεια ετεροζυγώσεως (LOH) για να εντοπίσει και να χαρακτηρίσει τα μεταγραφικά προγράμματα που οδηγούν τους διακριτούς γενετικούς υποκλώνους σε ένα δείγμα όγκου.

Επίσης στον νευρολογικό τομέα, η μέθοδος scRNA-seq πέτυχε, αποκαλύπτοντας την ετερογενή φύση των εγκεφαλικών κυττάρων που εμπλέκονται στη νόσο του Alzheimer και τα διαφορετικά αποτελέσματα που σχετίζονται με τα διαφορετικά πρότυπα γονιδιακής έκφρασης τους. Σε μια άλλη περίπτωση, αξιοποιήθηκε η αλληλούχιση μεμονωμένων κυττάρων για τον εντοπισμό μονών νουκλεοτιδικών παραλλαγών (SNVs) σε νευρωνικά κύτταρα, αποδεικνύοντας πώς μπορούν να χρησιμοποιηθούν σωματικές μεταλλάξεις για την ανακατασκευή της αναπτυξιακής γενεαλογίας των νευρώνων, οι οποίοι ζουν εδώ και δεκαετίες σε μεταμιτωτική κατάσταση συσσωρεύοντας μεταλλάξεις υπεύθυνες για τη δημιουργία ένθετων δέντρων γενεαλογίας και τη σχετική πολυκλωνική αρχιτεκτονική .

Ενώ, για δείγματα αίματος, ήπατος και καρδιάς, η εισαγωγή αναλύσεων τροχιάς παρείχε νέες γνώσεις σχετικά με τις διαδικασίες διαφοροποίησης, επιτρέποντας την ανίχνευση της τύχης των προγονικών κυττάρων που αποκαλύπτουν την πλαστικότητα του μεταγραφικού τους μέσω της

ταυτοποίησης νέων μεταβατικών κυτταρικών καταστάσεων. Ωστόσο, τα ρυθμιστικά δίκτυα που οδηγούν αυτές τις διαδικασίες είναι πιο περίπλοκα και χαρακτηρίζονται από συγχυτικούς παράγοντες. Μια αμερόληπτη προσέγγιση για την αποσαφήνιση τέτοιων κυκλωμάτων και των μεταβολών τους είναι οι μελέτες διαταραχών, οι οποίες, χρησιμοποιώντας τον τεράστιο παραλληλισμό των τεχνολογιών ενός κυττάρου που συγχωνεύονται με την επεξεργασία με διαμεσολάβηση CRISPR [20], επιτρέπουν την εξουδετέρωση πολλαπλών γονιδίων που παράγουν ταυτόχρονα διαφορετικές κυτταρικές αποκρίσεις χρήσιμες για την αποσαφήνιση της λειτουργίας πολλαπλών παραγόντων και των αλληλεπιδράσεών τους σε δεκάδες χιλιάδες κύτταρα. Για να επεκταθεί αυτή η εφαρμογή στην ανάλυση πολλαπλών άσχετων ατόμων, αναπτύχθηκαν νέες μέθοδοι. Εργαλεία, όπως το Demuxlet, καθορίζουν την ταυτότητα του δείγματος, χρησιμοποιώντας δεδομένα γονοτυπίας (SNPs), για να χαρακτηρίσουν τη διατομική παραλλαγή και τον ειδικό για τον κυτταρικό τύπο γενετικό έλεγχο της γονιδιακής έκφρασης. Ομοίως, μια άλλη μελέτη, χρησιμοποίησε δεδομένα SNP για τον χαρακτηρισμό μεταβολών της συν-έκφρασης γονιδίων, προωθώντας έναν νέο τρόπο εντοπισμού γενετικών παραλλαγών που επηρεάζουν τα ρυθμιστικά δίκτυα.

Ένα άλλο σημαντικό θέμα είναι η αποκατάσταση ζημιών, καθώς η καλύτερη κατανόηση αυτών των μηχανισμών θα μπορούσε να μας επιτρέψει να εντοπίσουμε τους παράγοντες που εμπλέκονται στην επιτυχία ή την αποτυχία τέτοιων διαδικασιών, προσφέροντας νέες συμβουλές στην ανάπτυξη καλύτερων διαγνωστικών εργαλείων και προγνωστικών βιοδεικτών.

ΚΕΦΑΛΑΙΟ 3: ΧΩΡΙΚΗ ΜΕΤΑΓΡΑΦΙΚΗ

3.1. Εισαγωγή

Παρά τη συνεχιζόμενη επιτυχία της τεχνολογίας scRNA-seq, υπάρχει ένα κρίσιμο πρακτικό εμπόδιο: η ανάγκη απελευθέρωσης βιώσιμων κυττάρων από ολόκληρο τον ιστό χωρίς να προκαλείται στρες, κυτταρικός θάνατος και/ή κυτταρική συσσώρευση. Οι ανοσολόγοι έχουν ίσως επωφεληθεί περισσότερο από την scRNA-seq επειδή πολλά ανοσοκύτταρα (ιδιαίτερα τα T και B λεμφοκύτταρα) δεν είναι πάντα αγκυροβολημένα στους ιστούς και επομένως είναι σχετικά εύκολο να απομονωθούν από το κυκλοφορούν αίμα, τα λεμφοειδή όργανα, τον περιφερικό ιστό και ακόμη και τους όγκους. Αντίθετα, πολλοί άλλοι τύποι κυττάρων, π.χ. οι νευρώνες στον εγκέφαλο, παραμένουν λιγότερο επιδεκτικοί στην ανάλυση scRNA-seq, απαιτώντας εξειδικευμένα πρωτόκολλα διάστασης ιστών για την ανάκτησή τους. Ως εκ τούτου, καθαρά από τεχνολογική άποψη, υπήρξε μια ώθηση για τη διεξαγωγή μεταγραφικής σε άθικτο ιστό. Δεδομένου ότι οι χωρικές πληροφορίες διατηρούνται με τη μελέτη του άθικτου ιστού, αυτές οι μέθοδοι αναφέρονται ως χωρική μεταγραφική.

Η θέση κάθε κυττάρου, σε σχέση με τους γείτονές του και τις μη κυτταρικές δομές, μπορεί να παρέχει χρήσιμες πληροφορίες για τον ορισμό του κυτταρικού φαινοτύπου, της κυτταρικής κατάστασης και τελικά της λειτουργίας των κυττάρων και των ιστών. Η τοποθεσία μπορεί να καθορίσει τα σήματα στα οποία εκτίθενται τα κύτταρα. Ενώ τα ενδοκρινικά σήματα δρουν σε μακροσκοπικές κλίμακες, πολλοί άλλοι τύποι σημάτων δρουν σε γειτονικά κύτταρα μέσω αλληλεπιδράσεων κυττάρου-κυττάρου. Μια μορφή που μπορούν να λάβουν αυτά τα σήματα είναι οι υποδοχείς πρωτεϊνών που συνδέονται με την επιφάνεια των κυττάρων και τα ζεύγη συνδετών, το mRNA για το οποίο μπορεί να ανιχνευθεί με μεταγραφική. Ως εκ τούτου, ένας πρωταρχικός οδηγός για ταχείες εξελίξεις στη χωρική μεταγραφική είναι ο ισχυρισμός ότι το πλαίσιο των ιστών βοηθά στην αξιολόγηση της κυτταρικής βιολογίας.

Επιπλέον, γίνεται όλο και πιο εμφανές ότι ο υπο-κυτταρικός εντοπισμός των mRNA ποικίλλει ανάλογα με τη γονιδιακή λειτουργία, ρυθμίζοντας για παράδειγμα εάν ένα πρωτεϊνικό προϊόν παράγεται και διακινείται στα κύτταρα. Αυτό είναι ένα κοινό φαινόμενο, που επηρεάζει περίπου το 70% των μεταγραφικών. Τις τελευταίες δεκαετίες, αυτά τα συμπεράσματα έγιναν από

στοχευμένες παρακολουθήσεις συγκεκριμένων mRNAs, αλλά είναι πέρα από τις τρέχουσες δυνατότητες της scRNA-seq ανάλυσης.

Αν και οι τεχνολογίες για την καταμέτρηση και τη δημιουργία προφίλ μεταγραφών στον ιστό υπάρχουν εδώ και δεκαετίες, μόνο το 2021 η χωρική μεταγραφική ονομάστηκε «μέθοδος του έτους 2020» από τη Nature Methods. Το πεδίο συνεχίζει να αναπτύσσεται γρήγορα, καθοδηγούμενο από πολλούς παράγοντες, όπως το μειωμένο κόστος Αλληλούχισης Επόμενης Γενιάς (NGS) και πρωτοβουλίες όπως ο Άτλας Ανθρώπινων Κυττάρων (HCA). Κατά τη διάρκεια αυτής της φάσης ανάπτυξης, οι υπολογιστικές προσεγγίσεις για τη δημιουργία και την ανάλυση χωρικών μεταγραφικών δεδομένων εξελίσσονται και βελτιώνονται γρήγορα, όπως συνέβη για την scRNA-seq ανάλυση από το 2009. Ενώ οι τεχνικές scRNA-seq έχουν δει ευρεία υιοθέτηση, μεγάλο μέρος της τρέχουσας βιβλιογραφίας για την χωρική μεταγραφική είναι τεχνική και δεν προσανατολίζεται σε ερευνητές που δεν είναι εξοικειωμένοι με το πεδίο.

3.2. Χωρική μεταγραφική στην τρέχουσα βιοϊατρική έρευνα

Οι χωρικές μεταγραφικές τεχνικές [21] υπήρχαν για σχεδόν μια δεκαετία, αλλά μέχρι πρόσφατα μόνο στα ινστιτούτα όπου αναπτύχθηκαν. Είναι γεγονός ότι οι εμπορευματοποιημένες τεχνικές κατέστησαν τη χωρική μεταγραφική προσιτή.

Σε ευρύ επίπεδο, η χωρική μεταγραφική μπορεί να χρησιμοποιηθεί για να χαρακτηρίσει τη μεταγραφική διαμόρφωση και ρύθμιση στους ιστούς. Για παράδειγμα, στην αναπαραγωγική βιολογία, ένας στόχος ήταν να καθοριστούν μηχανισμοί που ρυθμίζουν την κυτταρική διαφοροποίηση στο ενδομήτριο, το βλεννογόνο στρώμα της μήτρας, καθώς μεγαλώνει και αποβάλλεται κατά την εμμηνόρροια. Η εφαρμογή των Visium [22] και scRNA-seq σε ανθρώπινα δείγματα ενδομητρίου αποκάλυψε πρότυπα γονιδιακής έκφρασης σε υγιείς ιστούς σε διαφορετικές καταστάσεις επιθηλιακών κυττάρων που βρίσκονται σε συγκεκριμένες περιοχές του ιστού.

Η απεικόνιση μεγάλου οπτικού πεδίου μπορεί επίσης να αποκαλύψει νέα δομικά χαρακτηριστικά σε άρρωστους ιστούς. Πρόσφατες χωρικές πρωτεομικές μελέτες, που στοχεύουν 40-50

πρωτεΐνες σε ανθρώπινους όγκους, αποκάλυψαν κλίσεις ιστολογικών χαρακτηριστικών, όπως ο βαθμός και η μορφολογία που συνέπεσαν με κλίσεις πρωτεϊνών στον καρκίνο του παχέος εντέρου και ανοσολογικές συσχετίσεις πλήρους ανταπόκρισης στη θεραπεία στον καρκίνο του μαστού HER2+ . Οι τεχνικές αυτού του τύπου έχουν επίσης αποκαλύψει γονιδιωματικά χαρακτηριστικά σε υγιή και νοσούντα ιστό, για παράδειγμα τη χωρική κατανομή κλώνων καρκίνου σε μοντέλα μετάστασης ποντικού και πρωτοπαθή καρκίνο του ανθρώπου. Έτσι, οι τεχνικές αυτές μπορούν να βελτιώσουν την κατανόησή για την αρχιτεκτονική των ιστών και τη μοριακή της υποστήριξη στην υγεία και τις ασθένειες.

Σε ένα λεπτότερο επίπεδο, η χωρική μεταγραφική μπορεί να αποκαλύψει γειτονιές ιστών και τοπικά χαρακτηριστικά που συμβάλλουν στην ασθένεια. Στην νευροεπιστήμη, τα πλεονεκτήματα της χωρικής μεταγραφικής είναι διπλάσια. Πρώτον, αφαιρεί την ανάγκη για διαχωρισμό ιστών ευαίσθητων νευρώνων. Δεύτερον, διατηρεί το χωρικό πλαίσιο των κυττάρων. Μία μελέτη χρησιμοποίησε χωρική μεταγραφική για τον προσδιορισμό γονιδιακών ενοτήτων που εκφράζονται στην τοπική γειτνίαση με πλάκες αμυλοειδούς στο μοντέλο της νόσου του Alzheimer των ποντικών. Χρησιμοποιώντας χωρική μεταγραφική και σε αντίθεση με προηγούμενες αναφορές, αυτή η μελέτη πρότεινε ότι η εγγύτητα σε πλάκες αμυλοειδούς προκάλεσε ρύθμιση γονιδιακής έκφρασης για φλεγμονή, ενδοκυττάρωση και λυσοσωμική αποικοδόμηση. Παρατήρησαν επίσης αλλαγές στα ολιγοδενδροκύτταρα, συμπεριλαμβανομένων των γονιδίων μυελίνωσης. Τέλος, αυτές οι μεταγραφικές αλλαγές παρατηρήθηκαν επίσης στον ανθρώπινο ιστό χρησιμοποιώντας μια διαφορετική στοχευμένη, υψηλής ανάλυσης χωρική μεταγραφική τεχνική, *in situ sequencing* (ISS) [23]. Σε αυτή την περίπτωση, το χωρικό πλαίσιο αποκάλυψε διαφορετική ρύθμιση των ανοσολογικών γονιδίων, ιδιαίτερα των γονιδίων συμπληρώματος στην περιοχή των αμυλοειδών πλακών, υποδηλώνοντας έναν νέο μηχανισμό ασθένειας.

Η χωρική μεταγραφική έχει χρησιμοποιηθεί για να μελετήσει λεπτομερώς μεμονωμένα κύτταρα και τύπους κυττάρων. Για παράδειγμα, μια πρόσφατη μελέτη για τον καρκίνο του μαστού στον άνθρωπο χρησιμοποίησε μια αμερόληπτη (δηλαδή χαρακτηρίζοντας την έκφραση κάθε γονιδίου στο γονιδίωμα) τεχνική, την Visium, καθώς και scRNA-seq. Οι συγγραφείς προσδιόρισαν

υπολογιστικά περιοχές ιστών που αντιπροσωπεύουν περιοχές κυττάρων όγκου αποκλειστικά από προφίλ γονιδιακής έκφρασης.

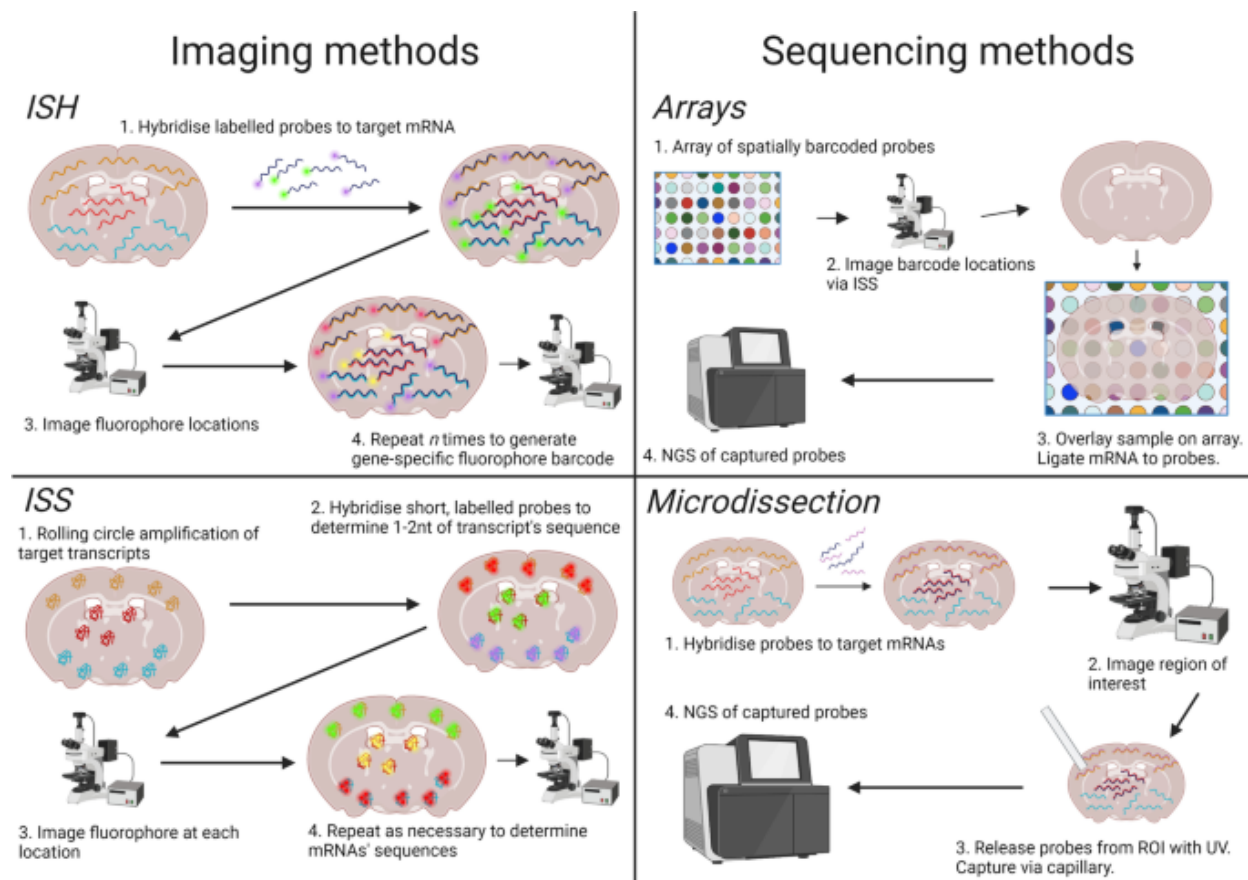
Συνολικά, η χωρική μεταγραφική είναι ευρέως εφαρμόσιμη σε ιστούς ανθρώπων και ποντικών τόσο σε σταθερή κατάσταση όσο και σε ασθένειες. Οι χωρικές τεχνικές μπορούν να αποκαλύψουν χαρακτηριστικά από το μοτίβο σε ολόκληρο τον ιστό έως τις κυτταρικές ίνες που προάγουν ασθένειες και ακόμη και τη βιολογία ενός κυττάρου. Μέχρι πρόσφατα, υπήρχαν μόνο λίγες εμπορικές πλατφόρμες χωρικής μεταγραφής. Όλο και περισσότερο, ένα ευρύτερο φάσμα τεχνολογιών με διαφορετικά τεχνικά θεμέλια και ερευνητικές δυνατότητες εισέρχονται στην αγορά.

3.3. Τεχνολογίες Χωρικής Μεταγραφικής

3.3.1. Τεχνολογίες που βασίζονται στην απεικόνιση

Η χωρική μεταγραφική με βάση την απεικόνιση χρησιμοποιεί *in situ* υβριδισμό mRNA με επισημασμένους συμπληρωματικούς ανιχνευτές, που ανιχνεύονται με μικροσκοπία για την ποσοτικοποίηση των μεταγραφών. Μια πρόσφατη ανασκόπηση προσδιορίζει τα πρώτα βήματα προς αυτές τις τεχνολογίες ως αρχές της δεκαετίας του 1960, με επισήμανση νουκλεϊνικών οξέων μέσω συμπληρωματικών ανιχνευτών το 1969, επισήμανση συγκεκριμένων ακολουθιών το 1973, και επισήμανση φθορισμού στα τέλη της δεκαετίας του '70. Αυτός ο *IN situ* υβριδισμός (ISH) [24] χρησιμοποιήθηκε για πρώτη φορά σε έναν ολόκληρο οργανισμό το 1989. Η μέθοδος ISH ήταν χρήσιμη για τη δημιουργία προφίλ προτύπων γονιδιακής έκφρασης σε τμήματα ιστών και ολόκληρους οργανισμούς, αλλά ήταν σε μεγάλο βαθμό ποιοτική. Μια ποσοτική μέθοδος για την απεικόνιση μεταγραφών με ετικέτα ανιχνευτή εμφανίστηκε το 1998, ο υβριδισμός φθορισμού ενός μορίου *in situ* (smFISH), στην οποία κάθε επισημασμένη μεταγραφή εμφανίζεται ως ένα μόνο σημείο μέσω μικροσκοπίας. Παραλλαγές αυτού έχουν διατεθεί στο εμπόριο εδώ και αρκετό καιρό, όπως η RNAscope. Η σύγχρονη χωρική μεταγραφική βασισμένη στην απεικόνιση βασίζεται στην smFISH [25]. Όπως και στις

πρόσφατες βιβλιογραφικές μελέτες, καθορίζουμε παρακάτω δύο κατηγορίες χωρικής μεταγραφικής με βάση την απεικόνιση: βασισμένη σε ISH (ISH – based) και βασισμένη σε ISS (ISS – based).



Εικόνα 2: Είδη Χωρικής Μεταγραφικής

Οι τρέχουσες μέθοδοι ISH περιλαμβάνουν τις seqFISH, seqFISH+ και MERFISH [26]. Και οι τρεις επεκτείνονται τη μέθοδο smFISH χρησιμοποιώντας πολλαπλούς γύρους υβριδισμού. Αντί για αντιστοιχία φθοροφόρου προς γονίδιο 1:1, κάτι που είναι αδύνατο για χιλιάδες γονίδια, η seqFISH χρησιμοποιεί έναν χρονικό γραμμωτό κώδικα φθοροφόρων στον οποίο οι ίδιοι ανιχνευτές υβριδοποιούνται σε διαφορετικούς γύρους υβριδισμού, αλλά κάθε φορά με διαφορετικό φθοροφόρο. Ο συνδυασμός και η σειρά των φθοροφόρων είναι ειδικός σε ένα γονίδιο. Αντίθετα, η μέθοδος seqFISH+ χρησιμοποιεί πρωτογενείς ανιχνευτές υβριδισμού ειδικών γονιδίων, ο καθένας με πολλαπλές τοποθεσίες σύνδεσης για υβριδισμό δευτερογενών ανιχνευτών με σήμανση φθοροφόρου. Σε κάθε διαδοχικό γύρο υβριδισμού, η επόμενη θέση

σύνδεσης υβριδοποιείται σε έναν ανιχνευτή και απεικονίζεται ολόκληρο το δείγμα, με τη ακολουθία των φθοροφόρων να ορίζει το είδος του mRNA.

Η μέθοδος seqFISH χρησιμοποιεί 4 φθοροφόρα. Η μέθοδος seqFISH + μπορεί να σχεδιάσει 10.000 γονίδια σε ένα μόνο κύτταρο χρησιμοποιώντας, αντί για 4-5 φθοροφόρα, 60 «ψευδοχρώματα» και επισημαίνοντας μόνο ένα κλάσμα μεταγραφών με φθοροφόρο σε κάθε γύρο υβριδισμού, αποφεύγοντας τον οπτικό συνωστισμό. Κάθε ψευδοχρώμα προέρχεται από συνδυασμό 20 μεμονωμένων γύρων υβριδισμού.

Η μέθοδος MERFISH χρησιμοποιεί επίσης διαδοχικό υβριδισμό, αλλά αντί για μια ακολουθία φθοροφόρων χρησιμοποιεί μια ακολουθία τοποθεσιών για δυαδικούς κωδικοποιημένους δευτερεύοντες ανιχνευτές: φθοροφόρο-επισημασμένο ή μη επισημασμένο. Κάθε γύρος υβριδισμού σε οποιαδήποτε τεχνική απεικόνισης υπόκειται σε κίνδυνο σφάλματος. Ο κίνδυνος σφάλματος για οποιαδήποτε μεταγραφή αυξάνεται εκθετικά με κάθε γύρο. Η δυαδική προσέγγιση της μεθόδου MERFISH είναι ανθεκτική στο σφάλμα, καθώς μειώνει την πιθανότητα ενός ασυμβίβαστου σφάλματος από έναν γύρο που εμποδίζει την αναγνώριση μιας μεταγραφής, επειδή εάν προσδιοριστεί μια απροσδόκητη ακολουθία, μπορεί πιο εύκολα να διορθωθεί σε μια αναμενόμενη ακολουθία.

Οι μέθοδοι ISS, αντί να χρησιμοποιούν ειδικούς ανιχνευτές γονιδίων, χρησιμοποιούν ανιχνευτές για να σχεδιάσουν 1-2 βάσεις προετοιμασμένων και ενισχυμένων μεταγραφών στον ιστό. Κάθε ακολουθία βάσης ή 2 βάσεων συνδέεται με διαφορετικό φθοροφόρο, επιτρέποντας την απεικόνιση και την καταγραφή που οδηγεί τελικά στην ταυτοποίηση κάθε μεταγραφής.

Συμπερασματικά, οι μέθοδοι ISH και ISS είναι χρήσιμες λόγω της υψηλής χωρικής τους ανάλυσης, η οποία είναι ικανή να περιγράφει το mRNA σε υποκυτταρικό επίπεδο. Οι μέθοδοι ISH ανιχνεύουν το mRNA πιο αποτελεσματικά από τις μεθόδους ISS. Ωστόσο, οι μέθοδοι ISS μπορούν να εξετάσουν μεγαλύτερες περιοχές ιστών. Τέλος, και οι δύο τύποι μεθόδων υφίστανται παρόμοιους τεχνικούς περιορισμούς, ιδίως την απαίτηση για πολλές ώρες έως ημέρες χρόνου απεικόνισης σε μικροσκόπιο, δημιουργώντας έτσι terabytes δεδομένων.

3.3.2. Τεχνολογίες που βασίζονται σε αλληλούχιση

Οι τεχνολογίες αυτές συλλαμβάνουν mRNAs από τον ιστό, συνθέτουν cDNAs και στη συνέχεια μετρούν συγκεκριμένες αλληλουχίες γονιδίων μέσω NGS. Είναι σημαντικό ότι οι πληροφορίες θέσης διατηρούνται στο σημείο της σύλληψης του mRNA. Μια πρώιμη τεχνολογία προς τη συγκεκριμένη κατεύθυνση ήταν η τεχνολογία LCM [27], μέσω της οποίας επεξεργάστηκαν συγκεκριμένες περιοχές ιστών για μεταγραφικό προφίλ μέσω μικροσυστοιχιών.

Η μέθοδος αυτή ακολουθήθηκε από την Tomo-seq, στην οποία ο ιστός κρυστομήθηκε, με κάθε τμήμα να υποβάλλεται σε RNA-seq ανάλυση. Οι σύγχρονες τεχνικές μικροδιατομής περιλαμβάνουν την GeoMx DSP [28] με μεταβλητή χωρική ανάλυση σε μεμονωμένα κύτταρα, στα οποία τα γονίδια είναι γραμμοκωδικοποιημένα με ανιχνευτές γονιδίων. Όταν το υπεριώδες φως φωτίζει τον ιστό, οι ανιχνευτές απελευθερώνονται και αλληλουχίζονται. Συνολικά, οι τεχνικές που βασίζονται σε μικροδιατομή παρέχουν ένα χρήσιμο πρώτο σύνολο τεχνικών για τη δημιουργία προφίλ αμερόληπτων χωρικών μεταγραφών, αλλά περιορίζονται από τη χωρική τους ανάλυση, την υποβάθμιση των mRNA όταν η LCM χρησιμοποιείται για μικροδιατομή και από την ανάγκη επεξεργασίας πολλών δειγμάτων για αλληλούχιση.

Οι μέθοδοι συστοιχίας έχουν διάφορα πλεονεκτήματα και μειονεκτήματα σε σύγκριση με τις μεθόδους ISH και ISS. Συχνά παρουσιάζουν μεγαλύτερα τμήματα ιστών. Χρησιμοποιώντας NGS αντί για μικροσκοπία, αποφεύγουν τους αγωγούς επεξεργασίας εικόνας. Επίσης, δεν στοχεύουν και μπορούν να σχεδιάσουν ολόκληρο το μεταγραφικό σώμα για κάθε οργανισμό που χρησιμοποιεί πολυαδενυλιωμένο mRNA. Ωστόσο, τα ποσοστά χωρικής ανάλυσης και ανάκτησης mRNA είναι χαμηλότερα σε σχέση με την περίπτωση της ISH. Τέλος, μεταγραφές από διαφορετικά κύτταρα μπορούν να συλληφθούν στο ίδιο σημείο, πράγμα που σημαίνει ότι απαιτούνται εξελιγμένες αναλύσεις για να προσδιοριστεί ποιοι τύποι κυττάρων ήταν παρόντες σε κάθε σημείο.

3.4. Βήματα Χωρικής Μεταγραφικής

3.4.1. Προεπεξεργασία χωρικών μεταγραφικών Δεδομένων

Μία από τις πρώτες διαδικασίες που πρέπει να εκτελεστούν είναι η προετοιμασία χωρικών μεταγραφικών δεδομένων για ανάλυση, γνωστή ως προεπεξεργασία. Αυτή συνήθως χρησιμεύει για τη μετατροπή ακατέργαστων δεδομένων απεικόνισης ή αλληλούχισης σε ένα μητρώο μεταγραφικών μετρήσεων ανά γονίδιο από χωρικές περιοχές σύλληψης, με συγκεκριμένες μεθόδους να απαιτούνται για δεδομένα που παράγονται από διαφορετικές τεχνικές.

Οι τυπικές μέθοδοι που βασίζονται σε ISH και ISH θα διαβάσουν μια ακολουθία ή έναν συγκεκριμένο γονιδιακό γραμμωτό κώδικα σε πολλαπλούς γύρους υβριδισμού. Απαιτούνται διάφορα βήματα για τη μετατροπή αυτών των εικόνων σε μήτρα γονιδιακού σημείου. Αρχικά, οι εικόνες φιλτράρονται για να αφαιρεθεί το φόντο και ο θόρυβος. Δεύτερον, οι εικόνες από διαφορετικούς γύρους υβριδισμού ευθυγραμμίζονται έτσι ώστε η ίδια θέση ή σημείο εικονοστοιχείων σε κάθε μία να αντιπροσωπεύει την ίδια μεταγραφή. Τρίτον, τα σήματα σε κάθε σημείο συνδυάζονται σε έναν γραμμωτό κώδικα, που μπορεί να χρησιμοποιηθεί για να ταιριάξει τα σημεία με τα γονίδια.

Ένα τελευταίο βήμα στην προεπεξεργασία των δεδομένων είναι η εφαρμογή στατιστικού μετασχηματισμού στο μητρώο για να ληφθούν υπόψη οι διαφορές στο ρυθμό σύλληψης mRNA σε ολόκληρο τον ιστό. Αυτό είναι ένα σημαντικό βήμα για τα δεδομένα που παράγονται μέσω όλων των τεχνικών, αλλά ειδικά εκείνων με χαμηλότερα ή μεταβλητά ποσοστά σύλληψης, όπως οι μέθοδοι που βασίζονται σε συστοιχίες. Στην scRNA-seq, η διαδικασία καταγραφής των διαφορών στη σύλληψη mRNA μεταξύ διαχωρισμένων κυττάρων ονομάζεται κανονικοποίηση και η ίδια ορολογία εφαρμόζεται στην χωρική μεταγραφική.

3.4.2. Προσδιορισμός χωρικών χαρακτηριστικών

Ένας αρχικός στόχος για πολλές αναλύσεις scRNA-seq είναι να καθοριστούν οι τύποι των κυττάρων. Ένα ανάλογο βήμα στην ανάλυση της χωρικής μεταγραφικής μπορεί να είναι ο προσδιορισμός χωρικών χαρακτηριστικών όπως ανατομικές και μικροανατομικές δομές. Για τον εντοπισμό δομών, οι υπάρχοντες αλγόριθμοι μπορούν να ομαδοποιήσουν μεταγραφικά παρόμοια σημεία ή κύτταρα με αμερόληπτο τρόπο για να αποκαλύψουν χωρικά πρότυπα γονιδιακής έκφρασης. Ωστόσο, έχουν αναπτυχθεί νεότερες μέθοδοι για την ειδική μόχλευση χωρικών δεδομένων και τον εντοπισμό χαρακτηριστικών.

Χαρακτηριστικό παράδειγμα αποτελεί το BayesSpace [29], ένα Bayesian στατιστικό εργαλείο που χρησιμοποιεί μια προσέγγιση ομαδοποίησης, αν και με προηγούμενες χωρικές πληροφορίες, για να ομαδοποιήσει μεταγραφικά παρόμοια, κοντινά σημεία. Επιπλέον, μπορεί να εκτελέσει βελτίωση της ανάλυσης με την ανακατανομή των αριθμών γονιδίων από ολόκληρα σημεία σε δεδομένα που βασίζονται σε συστοιχίες,. Ομοίως, το XFuse [30] συνδυάζει βοηθητικές εικόνες ιστολογίας με δεδομένα γονιδιακής έκφρασης μέσω βαθιάς σύντηξης δεδομένων χωρικών χαρακτηριστικών για να συμπεράνει την έκφραση γονιδίων μεταξύ σημείων σε μια συστοιχία.

Συνολικά, η επιλογή της μεθόδου θα εξαρτηθεί από τα διαθέσιμα δεδομένα; για μεθόδους βασισμένες σε συστοιχίες με χαμηλή ανάλυση, το BayesSpace μπορεί να είναι ευνοϊκό για την ενίσχυση της ανάλυσης, αλλά εάν υπάρχουν ιστολογικές εικόνες, τότε το εργαλείο όπως το stLearn θα είναι ισχυρό μέσω της ικανότητάς του να τις ενσωματώνει στην ανάλυση των χωρικών μεταγραφικών δεδομένων.

3.4.3. Αποσυνέλιξη

Ένα σημαντικό στοιχείο για τα δεδομένα που παράγονται με μεθόδους συστοιχίας με μη κυτταρική ανάλυση είναι ότι περισσότεροι από ένας τύποι κυττάρων μπορούν να συμβάλουν σε κάθε σημείο. Η διαδικασία ταυτοποίησης και ποσοτικοποίησης της σχετικής συμβολής από κάθε τύπο κυττάρου σε ένα σημείο σύλληψης είναι γνωστή ως αποσυνέλιξη. Πολλά εργαλεία έχουν αναπτυχθεί για την εκτέλεση αυτής της εργασίας. Ένα πρώιμο παράδειγμα ήταν το NMFreg [31], μια προσέγγιση που αποσυνθέτει τα μεταγραφικά σημεία σε συνεισφορές από κυτταρικούς τύπους με μη αρνητική παραγοντοποίηση μήτρας. Το SPOTlight [31] χρησιμοποιεί επίσης μη αρνητική παραγοντοποίηση μήτρας. Ακολούθησε η ισχυρή αποσύνθεση τύπου κυττάρου (RCTD), η οποία χρησιμοποίησε ένα διαφορετικό στατιστικό μοντέλο για να εξηγήσει τον αριθμό των γονιδίων σε κάθε σημείο.

3.4.4. Καταλογισμός και χαρτογράφηση μεμονωμένων κυττάρων

Ορισμένες υπολογιστικές μέθοδοι στοχεύουν να συνδυάσουν την χωρική μεταγραφική και την μέθοδο scRNA-seq, όχι για λόγους αποσύνδεσης, αλλά για να συμπεράνουν σε ποιες περιπτώσεις τα γονίδια, ενώ δεν ανιχνεύονται, μπορεί στην πραγματικότητα να έχουν εκφραστεί. Αντίθετα, ορισμένες μέθοδοι ακολουθούν την αντίθετη προσέγγιση, χρησιμοποιώντας χωρικά σύνολα δεδομένων για να συμπεράνουν χωρικές αντιστοιχίσεις που προέρχονται από scRNA-seq, όπως για παράδειγμα στο Visium όπου δεν μπορούν να γίνουν μετρήσεις μονοκυττάρων.

Μια πρώιμη προσέγγιση ολοκλήρωσης ήταν το Seurat, το οποίο χαρτογραφεί τα μεταγράφημα ενός κυττάρου σε χωρικές συντεταγμένες.. Από τότε, έχουν επινοηθεί πολλές άλλες υπολογιστικές προσεγγίσεις. Πιο πρόσφατες προσεγγίσεις χαρτογράφησης περιλαμβάνουν το SpaOTsc [32], το οποίο βασίζεται σε ένα βέλτιστο μοντέλο μεταφοράς για τη χαρτογράφηση των μεταγραφών ενός κυττάρου σε χωρικά δεδομένα και περιλαμβάνει επίσης λειτουργίες για την εξαγωγή συμπερασμάτων αλληλεπίδρασης συνδέτη-υποδοχέα. Οι προσεγγίσεις καταλογισμού περιλαμβάνουν το gimVI, το οποίο μαθαίνει μια ευθυγράμμιση μεταξύ των δεδομένων scRNA-seq και των χωρικών μεταγραφών, που περιλαμβάνονται στο πακέτο ανάλυσης scVi, και το Tangram [33], το οποίο χρησιμοποιεί ένα βήμα χαρτογράφησης για να ενημερώσει τη διαδικασία καταλογισμού. Το gimVI και το Tangram είναι μοντέλα που βασίζονται σε βαθιά μάθηση, οπότε η επιλογή της μεθόδου μπορεί να εξαρτάται από το αν οι πόροι υπολογιστών GPU είναι διαθέσιμοι στους ερευνητές. Η χαρτογράφηση των μονοκύτταρων μεταγραφών πιθανότατα θα είναι χρήσιμη για δεδομένα μη μονοκύτταρων, όπως μη τμηματικά δεδομένα που βασίζονται σε απεικόνιση ή δεδομένα από οποιαδήποτε μέθοδο που βασίζεται σε αλληλουχία. Αντίθετα, η καταλογισμός θα είναι χρήσιμος για την εξαγωγή μη μετρημένων γονιδίων σε στοχευμένα χωρικά μεταγραφικά δεδομένα.

3.4.5. Συμπερασματολογία αλληλεπίδρασης κυττάρου-κυττάρου

Ένας κοινός στόχος στην ανάλυση των μεταγραφικών δεδομένων είναι να συναχθούν ενδοκυτταρικές αλληλεπιδράσεις με βάση την έκφραση των προσδεμάτων και των υποδοχέων. Συνήθως, ένα εργαλείο συμπεράσματος αλληλεπίδρασης κυττάρου-κυττάρου συνδυάζει μια βάση δεδομένων γονιδίων που κωδικοποιούν πρωτεΐνες που έχουν αποδειχθεί ότι συμμετέχουν σε διακυτταρικές αλληλεπιδράσεις, με έναν αλγόριθμο για να συμπεράνει την πιθανότητα

αλληλεπίδρασης από τα δεδομένα γονιδιακής έκφρασης. Υπάρχουν τώρα αρκετά παραδείγματα αυτών των εργαλείων, συμπεριλαμβανομένων των CellPhoneDB V.2. 0, iCellNet, CellChat και SingleCellSignalR [34]. Ενώ είναι αποτελεσματικά για μονοκύτταρα και ορισμένα χωρικά δεδομένα, αυτά τα εργαλεία δεν μπορούν να αξιοποιήσουν χωρικά δεδομένα όταν είναι διαθέσιμα. Πρόσφατα, έχουν αναπτυχθεί αρκετές τεχνικές για το σκοπό αυτό, συμπεριλαμβανομένων των SpaOTsc, cell2cell, MISTy και CellPhoneDB V.3.0. Ορισμένες από αυτές τις τεχνικές, μπορούν επίσης να συμπεράνουν διαφορική έκφραση γονιδίων με εγγύτητα σε ένα κύτταρο αποστολής σήματος.

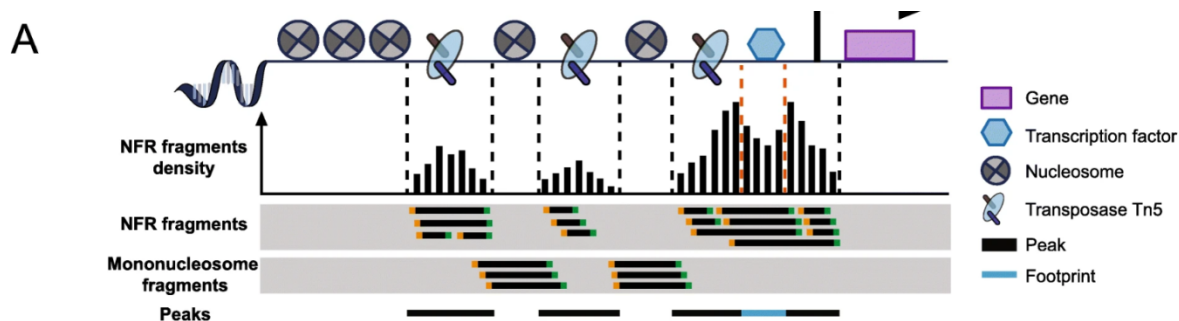
Συμπερασματικά, απαιτείται μια σειρά τεχνικών για να συναχθούν βιολογικές διεργασίες από τη χωρική μεταγραφική: μέθοδοι ειδικές για την τεχνική για την προετοιμασία δεδομένων, για τον εντοπισμό περιοχών που αντιστοιχούν σε μεμονωμένα κύτταρα και περιοχές ιστών μόνο από μεταγραφικά, και μέθοδοι για να συναχθούν γονίδια που μπορεί να μην έχουν διαμορφωθεί από τη χωρική μεταγραφική. Είναι αδιαμφισβήτητο ότι στο μέλλον θα αναπτυχθούν περισσότερο εξελιγμένες τεχνικές με ακόμη μεγαλύτερη αποτελεσματικότητα.

ΚΕΦΑΛΑΙΟ 4: ATAC-seq

4.1. Εισαγωγή

Πρόσφατες μελέτες γονιδιακής ρύθμισης επικεντρώθηκαν στην επιγενετική και μέσω της προόδου των τεχνολογιών αλληλούχισης υψηλής απόδοσης, έχουν αναπτυχθεί διάφορες δοκιμές για την αποκρυπτογράφηση του επιγενετικού τοπίου. Αυτές περιλαμβάνουν τη ATAC-seq [35] (Assay of Transposase Accessible Chromatin sequencing), την αλληλούχιση υπερευαίσθητων θέσεων DNase I (DNase-seq) και την υποβοηθούμενη από φορμαλδεΰδη απομόνωση αλληλουχίας ρυθμιστικών στοιχείων (FAIRE-seq).

Μεταξύ των δοκιμών που έχουν σχεδιαστεί για την ανίχνευση της προσβασιμότητας της χρωματοίνης, η ATAC-seq έχει αποκτήσει ιδιαίτερη δημοτικότητα από τότε που περιγράφηκε για πρώτη φορά το 2013. Μια εκθετική αύξηση των επιμελημένων συνόλων δεδομένων και δημοσιεύσεων σχετικά με την ATAC-seq δείχνει την αξία της σε ένα ευρύ φάσμα βιολογικών ερωτήσεων, όπως η απεικόνιση σε υγιείς τύπους ιστών και κυττάρων θηλαστικών, μελετώντας τις αλλαγές προσβασιμότητας μεταξύ της φυσιολογικής αιματοποίησης και της λευχαιμίας, καθώς και την κατάσταση χρωματίνης στους ασθενείς με σχιζοφρένεια. Ένα σχηματικό διάγραμμα αυτής της τεχνολογίας παρουσιάζεται στην επόμενη εικόνα. Εν συντομία, η ATAC-seq ενσωματώνει μια γενετικά τροποποιημένη υπερκινητική τρανσποζάση Tn5 που ταυτόχρονα αφαιρεί την ανοιχτή χρωματίνη και συνδέει προσαρμογείς αλληλουχίας υψηλής απόδοσης σε αυτές τις περιοχές. Στη συνέχεια, πραγματοποιείται αλληλούχιση ζευγαρωμένου άκρου για να προκύψουν υψηλότεροι μοναδικοί ρυθμοί ευθυγράμμισης αυτών των περιοχών.

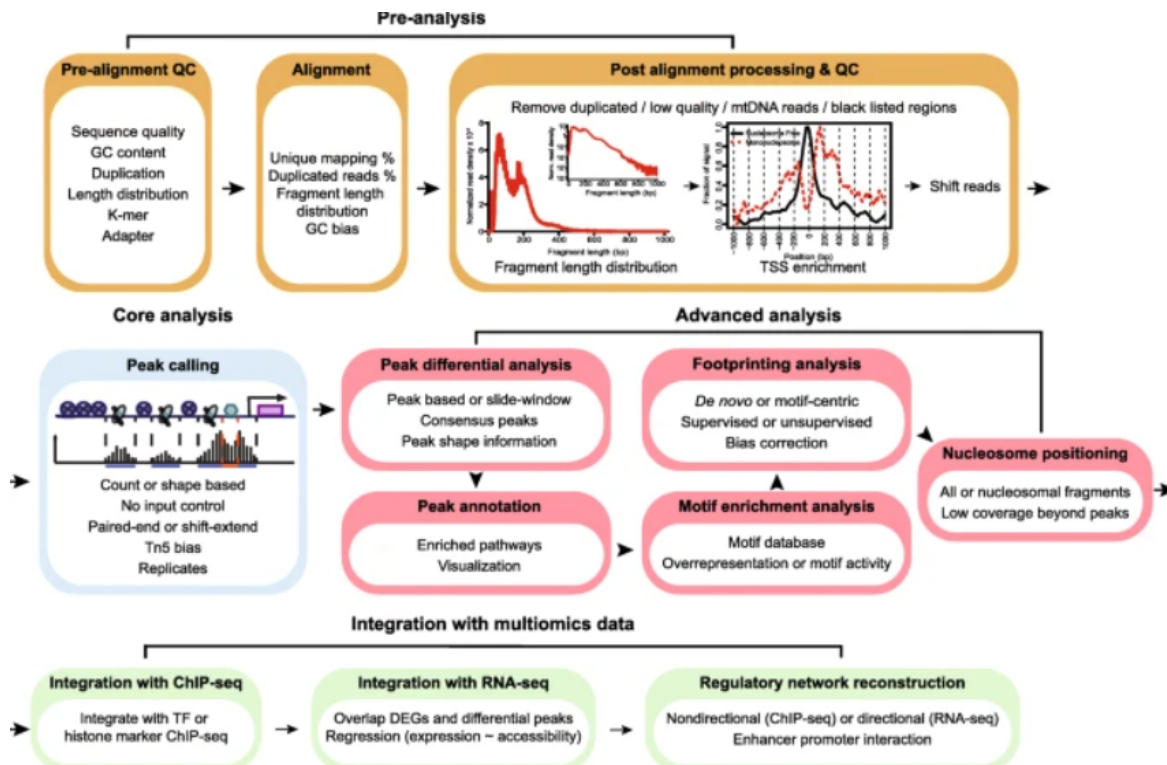


Εικόνα 3: Βασική μεθοδολογία ATAC - seq

Η υπερδραστηριότητα της τρανσποζάσης Tn5 καθιστά το πρωτόκολλο ATAC-seq μια απλή, χρονικά αποδοτική μέθοδο που απαιτεί 500-50.000 κύτταρα. Η ευαισθησία και η ειδικότητα είναι συγκρίσιμες με την DNase-seq αλλά ανώτερες από την FAIRE-seq όπου και οι δύο μέθοδοι απαιτούν εκατομμύρια κύτταρα ως υλικό εισόδου. Επειδή η ATAC-seq δεν περιλαμβάνει αυστηρή επιλογή μεγέθους κατά την προετοιμασία, μπορεί επίσης να εντοπίσει θέσεις νουκλεοσωμάτων χρησιμοποιώντας θραύσματα που αντιπροσωπεύουν μονομερές νουκλεοσωμάτων και πολυμερές. Πρόσφατα, έχει περιγραφεί η μονοκύτταρη ATAC-seq (scATAC-seq), χρησιμοποιώντας προσεγγίσεις κυτταρικής ταξινόμησης (FACS) ενεργοποιημένες με φθορισμό. Η ScATAC-seq μπορεί να εφαρμοστεί σε πολλαπλές καταστάσεις, συμπεριλαμβανομένων κλινικών δειγμάτων και αναπτυξιακής βιολογίας για τη μελέτη των ετερογενών κυτταρικών πληθυσμών σε ανάλυση ενός κυττάρου.

Παρά την απλότητα και την ευρωστία της ATAC-seq, υπάρχει ένα σημαντικό εμπόδιο καθώς υπάρχουν λίγα εργαλεία βιοπληροφορικής ανάλυσης που αναπτύχθηκαν ειδικά για τα δεδομένα ATAC-seq. Τα εργαλεία ανάλυσης που χρησιμοποιούνται στις ChIP-seq και DNase-seq έχουν εφαρμοστεί στην ATAC-seq υποθέτοντας παρόμοια χαρακτηριστικά δεδομένων. Ωστόσο, αυτή η υπόθεση δεν έχει αξιολογηθεί συστηματικά.

Στόχος του κεφαλαίου αυτού είναι να παρατεθεί μια συνοπτική, αλλά πλήρης ανάλυση όσο αφορά τη μέθοδο αυτή.



Εικόνα 4: Βήματα ATAC - seq

4.2. Αρχή και διαδικασίες της ATAC-seq

Η ATAC-seq είναι μια καινοτόμος επιγενετική τεχνολογία [36], η οποία είναι μια μέθοδος για τη χαρτογράφηση της προσβασιμότητας της χρωματίνης με ανίχνευση υπερκινητικής τρانشποζάσης Tn5 σε αλληλουχία DNA σε επίπεδο γονιδιώματος. Το τρانشποζόνιο DNA είναι ένα φαινόμενο που μεταφέρει αλληλουχία DNA από μια περιοχή χρωμοσώματος σε άλλη. Το τρانشποζόνιο DNA απαιτεί η χρωματίνη στη θέση εισαγωγής να είναι ανοιχτή και η τρانشποζάση που φέρει γνωστές ετικέτες αλληλουχίας DNA πρέπει να προστεθεί τεχνητά στον πυρήνα και στη συνέχεια η ανοιχτή χρωματίνη μπορεί να αναγνωριστεί χρησιμοποιώντας ετικέτες γνωστών αλληλουχιών για την κατασκευή μιας βιβλιοθήκης για αλληλούχιση. Επί του παρόντος, η πιο συχνά χρησιμοποιούμενη τρانشποζάση είναι η τρانشποζάση Tn5, η οποία μπορεί να μεταφερθεί στην προσβάσιμη χρωματίνη πιο συχνά από ό, τι στην απρόσιτη χρωματίνη. Η τρانشποζάση Tn5 δρα ως ανιχνευτής για τη μέτρηση της προσβασιμότητας της

χρωματίνης σε επίπεδο γονιδιώματος μέσω του μηχανισμού «κοπής και επικόλλησης» και το τρανσποζόνιο μπορεί ταυτόχρονα να κατακερματίζει και να επισημάνει τις μη προστατευμένες περιοχές του DNA με προσαρμογείς αλληλουχίας.

Η κατασκευή της βιβλιοθήκης ATAC-seq αποτελείται από τρία στάδια: προετοιμασία πυρήνων, μεταφορά και ενίσχυση. Πρώτον, οι ιστοί ή τα κύτταρα για εξέταση βρίσκονται σε άθικτα, ομοιογενή μεμονωμένα κύτταρα, τα οποία στη συνέχεια επωάζονται στο ρυθμιστικό διάλυμα για τη δημιουργία ακατέργαστων πυρήνων. Δεύτερον, οι πυρήνες επωάζονται στο μίγμα αντίδρασης για να αποδώσουν τμήμα του DNA. Τέλος, το μετατοπισμένο DNA ενισχύεται για να δημιουργήσει βιβλιοθήκες για αλληλούχιση. Η αντίδραση του μεταφερόμενου ενζύμου στη χρωματίνη του δείγματος είναι το βασικό βήμα του πειράματος ATAC.

Ο ποιοτικός έλεγχος της βιβλιοθήκης ATAC-seq θα πρέπει να πραγματοποιείται πριν από την αλληλούχιση για να διασφαλιστεί ότι πληρούνται τα κριτήρια αλληλουχίας. Μετά την αλληλούχιση της βιβλιοθήκης, οι ακατέργαστες αναγνώσεις συλλέγονται μέσω της αλληλούχισης της ειδικευμένης βιβλιοθήκης. Μετά το φιλτράρισμα των δεδομένων μέσω της αξιολόγησης της ποιότητας των δεδομένων αλληλουχίας, οι καθαρές αναγνώσεις λαμβάνονται περαιτέρω με την αξιολόγηση της ποιότητας αλληλουχίας και την περίληψη της παραγωγής δεδομένων. Μετά την αφαίρεση των αλληλουχιών προσαρμογέα, η υψηλής ποιότητας ανάγνωση περίπου 150 νουκλεοτιδίων (nts) σε μήκος επεξεργάζεται για περαιτέρω ανάλυση. Οι αναγνώσεις χαρτογραφούνται στο γονιδίωμα αναφοράς και στις προσβάσιμες περιοχές χρωματίνης, όπως προωθητές, ενισχυτές και μονωτές. Μια σειρά λεπτομερών αναλύσεων μπορεί να διεξαχθεί περαιτέρω, όπως η εξακρίβωση της κατανομής των αναγνώσεων σε ολόκληρο το γονιδίωμα, ο προσδιορισμός της κατανομής του μήκους κορυφής, η λειτουργική ανάλυση γονιδίων με αναγνωρισμένες κορυφές, η κατανομή κορυφών σε λειτουργικά στοιχεία γονιδίων και η διαφορική ανάλυση κορυφών μεταξύ δειγμάτων.

4.3. Πλεονεκτήματα και περιορισμοί

Η μέθοδος ATAC-seq αναπτύχθηκε αρχικά ως εναλλακτική λύση ή συμπλήρωμα στην αλληλούχιση ευαίσθητων σε μικροκοκκική νουκλεάση θέσεων (MNase-seq), την

υποβοηθούμενη από φορμαλδεΰδη απομόνωση ρυθμιστικών στοιχείων (FAIRE-seq) και την αλληλούχιση υπερευαίσθητων θέσεων δεοξυριβονουκλεάσης I (DNase-seq). Στην MNase-seq και την DNase-seq, όταν μειώνεται η συσσωμάτωση DNA και ιστόνης, το μη προστατευμένο DNA εκτίθεται και κόβεται από ένζυμα DNA. Οι προσβάσιμες περιοχές χρωματίνης αναγνωρίζονται με την αλληλούχιση των διασπασμένων τμημάτων DNA και τη σύγκριση της αλληλουχίας στο γονιδίωμα αναφοράς. Τα μειονεκτήματα αυτών των δύο μεθόδων είναι η χρονοβόρα και κακή επαναληψιμότητα. Η FAIRE-seq χρησιμοποιεί ακινητοποίηση φορμαλδεΰδης, εκχύλιση χλωροφόρμου φαινόλης και διαχωρισμό για να αποκτήσει το εκτεθειμένο DNA. Στην περίπτωση αυτή, η αναλογία σήματος προς θόρυβο αλληλουχίας είναι χαμηλή.

Επιπλέον, η ATAC-seq παρουσιάζει πολλά πλεονεκτήματα [36]: Πρώτον, η μέθοδος μπορεί να μειώσει τον πειραματικό χρόνο σε 2-3 ώρες για να επιτύχει κατακερματισμό DNA χρησιμοποιώντας μια απλή ενζυματική αντίδραση, η οποία αποφεύγει τον κουραστικό συμβατικό κατακερματισμό DNA και την αντίδραση σύνδεσης προσαρμογέα. Από την άλλη πλευρά, συνήθως χρειάζονται 2-3 ημέρες για την προετοιμασία των πειραμάτων των DNase-seq και MNase-seq και 3-4 ημέρες για το πείραμα της FAIRE-seq. Δεύτερον, η απλοποιημένη πειραματική διαδικασία μειώνει τη διάρκεια της προετοιμασίας του δείγματος και μειώνει την πιθανότητα σφαλμάτων, γεγονός που βελτιώνει σημαντικά τον επιτυχή ρυθμό και την επαναληψιμότητα ενός πειράματος. Τρίτον, το μέγεθος του δείγματος μειώνεται κατά τουλάχιστον 1000 φορές. Όταν η συλλογή δειγμάτων είναι δύσκολη, αυτό το πλεονέκτημα είναι ιδιαίτερα εμφανές. Τέταρτον, η ATAC-seq μπορεί να χρησιμοποιήσει τεχνολογία αλληλουχίας ζευγαρωμένου άκρου για να χαρτογραφήσει την τοποθέτηση των νουκλεοσωμάτων. Η αλληλουχία ζευγαρωμένου άκρου μπορεί να αλληλουχήσει και τα δύο άκρα του τμήματος DNA, καθιστώντας την ευθυγράμμιση της χαρτογράφησης ανάγνωσης σε επαναλαμβανόμενες περιοχές του γονιδιώματος πιο ακριβή.

Υπάρχουν επίσης ορισμένοι περιορισμοί της τεχνολογίας ATAC-seq. Πρώτον, η τρανσποζάση Tn5 ταυτόχρονα θραύει και επισημαίνει μη προστατευμένες περιοχές DNA με προσαρμογείς αλληλουχίας μέσω του μηχανισμού "κοπής και επικόλλησης". Οι αρμοί προσαρμογέα και στα δύο άκρα κάθε θραύσματος DNA είναι τυχαίοι, γεγονός που οδηγεί σε πιθανότητα 50% ότι οι προσαρμογείς και στα δύο άκρα ενός θραύσματος είναι οι ίδιοι, δημιουργώντας μισά άχρηστα

θραύσματα για εμπλουτισμό, ενίσχυση και αλληλούχιση. Δεύτερον, μελέτες έχουν δείξει ότι το DNA χωρίς νουκλεοσώματα και παράγοντες μεταγραφής είναι ευκολότερο να διασπαστεί από την τρανσποζάση Tn5. Επιπλέον, η τρανσποζάση Tn5 τείνει να δεσμεύεται και να διασπάται σε περιοχές δέσμευσης παράγοντα μεταγραφής, γεγονός που έχει ως αποτέλεσμα την απώλεια μέρους των πληροφοριών παράγοντα μεταγραφής. Όλα αυτά τα μειονεκτήματα καθιστούν δύσκολο η ATAC-seq να ανιχνεύσει το αποτύπωμα των παραγόντων μεταγραφής, το οποίο μπορεί να χρησιμοποιηθεί για τον εντοπισμό πιθανών μοτίβων δέσμευσης των παραγόντων μεταγραφής. Τρίτον, λόγω της παρουσίας μιτοχονδριακού DNA, τα δεδομένα που λαμβάνονται από την ATAC-seq περιέχουν αναπόφευκτα ορισμένες μιτοχονδριακές αναγνώσεις. Ανάλογα με τον τύπο των κυττάρων, τα δεδομένα ATAC-seq μπορεί να περιέχουν 20-80% των μιτοχονδριακών αλληλουχιών διαβάσει .

4.4. Βελτίωση της ATAC-seq

Από τότε που αναπτύχθηκε για πρώτη φορά η μέθοδος ATAC-seq, έχει βελτιωθεί προκειμένου να προσαρμοστεί η ευρύτερη χρήση στην έρευνα. Η μονοκύτταρη ATAC-seq (scATAC-seq) παρέχει την πρώτη διορατική εξέταση της μεταβλητότητας κυττάρου προς κύτταρο στην οργάνωση χρωματίνης, η οποία μπορεί να επιτευχθεί με προγραμματιζόμενη μικρορευστική συσκευή ή συνδυαστικό σύστημα κυτταρικής ευρετηρίασης. Η scATAC-seq [37] μπορεί να χρησιμοποιηθεί ως όχημα σε όλο το γονιδίωμα για τη χαρτογράφηση της προσβασιμότητας της χρωματίνης σε όλους τους συγκεκριμένους κυτταρικούς τύπους ενός οργανισμού. Επειδή εξακολουθεί να είναι ασαφές πόσες περιοχές ανοιχτής χρωματίνης υπάρχουν σε ένα μόνο κύτταρο και πώς η προσβασιμότητα της χρωματίνης διαφέρει μεταξύ των δύο αλληλόμορφων σε ένα μεμονωμένο κύτταρο, το αν η scATAC-seq καταγράφει ένα περιορισμένο υποσύνολο ανοιχτών θέσεων χρωματίνης σε μεμονωμένα κύτταρα παραμένει ασαφές.

Επιπλέον, το Omni-ATAC-seq [38] είναι ένα άλλο βελτιωμένο πρωτόκολλο ATAC-seq για την ανίχνευση της προσβασιμότητας της χρωματίνης. Με βάση το πρότυπο πρωτόκολλο ATAC-seq, το Omni-ATAC-seq προσθέτει ένα βήμα καθαρισμού χρησιμοποιώντας μεθόδους για την απομάκρυνση των μιτοχονδρίων. Το Omni-ATAC-seq χρησιμοποιεί επίσης αλατούχο διάλυμα με φωσφορικό ρυθμιστικό διάλυμα (PBS) στην αντίδραση μεταφοράς για να αυξήσει τη σχέση

σήματος προς φόντο και να μειώσει το φόντο. Έτσι, το Omni-ATAC-seq εξαλείφει τις μιτοχονδριακές παρεμβολές και μειώνει τον θόρυβο του περιβάλλοντος για να αποκτήσει δεδομένα υψηλής ποιότητας προσβασιμότητας χρωματίνης. Επιπλέον, το πρότυπο ATAC-seq απαιτεί την αντίδραση μεταφοράς να εκτελείται σε φρέσκα κύτταρα και αργά ψυχθέντα κρυοσυντηρημένα κύτταρα. Το πρωτόκολλο Omni-ATAC μπορεί να δημιουργήσει προφίλ προσβασιμότητας χρωματίνης υψηλής ποιότητας από κλινικά σχετικούς κατεψυγμένους ιστούς, όπως εγκεφάλους. Επιπλέον, η ATAC-seq αποκτά τις πληροφορίες της προσβάσιμης χρωματοίνης διασπώντας τα κύτταρα, οπότε δεν μπορεί να περιγράψει την τρισδιάστατη δομή αυτών των προσβάσιμων γονιδιωματικών περιοχών.

4.5. Εφαρμογές

4.5.1. Χαρτογράφηση του προσβάσιμου τοπίου χρωματίνης

Η χαρτογράφηση του προσβάσιμου τοπίου χρωματίνης μπορεί να λάβει πληροφορίες για χωρικές αλλαγές στις δομές της χρωματίνης και παράγοντες μεταγραφής που σχετίζονται με την έκφραση γονιδίων. Αυτές οι πληροφορίες μπορούν να αποκαλύψουν το δίκτυο των σχετικών παραγόντων μεταγραφής και τους μηχανισμούς της δομικής ρύθμισης της χρωματίνης που διέπει τα προγράμματα γονιδιακής έκφρασης. Για παράδειγμα, στο ανθρώπινο ανοσοποιητικό σύστημα, ο προσβάσιμος χάρτης χρωματίνης των πρωτογενών ανοσοκυττάρων έχει ταυτοποιηθεί με τη χρήση ATAC-seq. Μια σημαντική αλλαγή της προσβασιμότητας της χρωματίνης έχει εντοπιστεί σε περιοχές κοντά σε γονίδια που σχετίζονται με την ενεργοποίηση των Β κυττάρων, ειδικά σε ασθενείς με συστηματικό ερυθηματώδη λύκο.

Στην αναπτυξιακή βιολογία, οι περιοχές ανοιχτής χρωματίνης έχουν χαρτογραφηθεί χρησιμοποιώντας ATAC-seq. Επιπλέον, στο νευρικό σύστημα, η επαγόμενη ενεργοποίηση των νευρώνων οδηγεί σε στιγμιαίες αλλαγές στη δομή της χρωματίνης..

Συνοπτικά, η εφαρμογή της ATAC-seq έχει δημιουργήσει ολοκληρωμένα προσβάσιμα τοπία χρωματίνης διαφόρων τύπων κυττάρων σε διαφορετικούς ιστούς και όργανα, γεγονός που έχει παράσχει πολύτιμες γνώσεις σχετικά με την πολυπλοκότητα της μεταγραφής των γονιδίων.

4.5.2. ATAC-seq στην εμβρυϊκή ανάπτυξη

Ο επαναπρογραμματισμός της χρωματίνης εμφανίζεται ενεργά κατά την πρώιμη εμβρυϊκή ανάπτυξη. Μελέτες έχουν δείξει ότι όταν συμβαίνει επαναπρογραμματισμός χρωματίνης, ρυθμιστικοί παράγοντες που σχετίζονται με τη μεταγραφή γονιδίων και τον ανασυνδυασμό DNA προσλαμβάνονται από χρωματίνη και ταυτόχρονα μεταβάλλεται η σταθερότητα του νουκλεοσώματος. Κατά τη διάρκεια της ενεργοποίησης του γονιδίου ζυγώτη, η δραστηριότητα της ανοικτής χρωματίνης αυξάνεται, παράλληλα με τις δραστηριότητες των -ρυθμιστικών παραγόντων. Η ATAC-seq έχει χρησιμοποιηθεί μαζί με την τεχνολογία CRISP/Cas-9, για την ανίχνευση εμβρύων ποντικών.

Επιπλέον, για τη μελέτη της ενεργοποίησης του εμβρυϊκού γονιδιώματος, η ATAC-seq έχει χρησιμοποιηθεί για την ανίχνευση αλληλουχιών μεταγραφών. Αυτές οι μελέτες έχουν δείξει ότι πολλαπλά αντίγραφα του DUX4 ενεργοποιούνται από ενδογενή γονίδια KDM4E και ZSCAN4 που εκφράζονται μόνο σε στάδιο διάσπασης ανθρώπινων εμβρύων.

4.5.3. Ανίχνευση καρκίνου

Η ATAC-seq είναι ιδιαίτερα εφαρμόσιμη για τη σύλληψη της ιστοειδικής δραστικότητας χρωματίνης των ρυθμιστικών περιοχών σε όγκους. Οι επαναλαμβανόμενες μεταλλάξεις στα γονίδια RAD21 και STAG2, τα οποία κωδικοποιούν το σύμπλεγμα συνοχής χρωμοσωμάτων, έχουν αποδειχθεί ότι είναι βασικά στοιχεία στον σχηματισμό κακοήθειας στην οξεία μυελογενή λευχαιμία (AML). Μελέτες έχουν δείξει ότι η μεταλλαγμένη συνεκτίνη μπορεί να αυξήσει την προσβασιμότητα της χρωματίνης στις θέσεις σύνδεσης για παράγοντες μεταγραφής όπως ERG, GATA2 και RUNX1, όπως ανιχνεύεται με τη χρήση ATAC-seq.

Επιπλέον, οι μεταλλάξεις ARID1A εμφανίζονται συνήθως σε πολλά είδη όγκων, όπως το μελάνωμα, το γλοιοβλάστωμα και άλλες ανθρώπινες κακοήθειες, και οι μεταλλάξεις ARID1B βρίσκονται συνήθως στο νευροβλάστωμα, στο ηπατοκυτταρικό καρκίνωμα και στο διηθητικό καρκίνωμα του μαστού. Μελέτες έχουν δείξει ότι οι μεταλλάξεις του συμπλέγματος ARID1A

και ARID1B συσχετίζονται συχνά με την ογκογένεση μεταβολών για τη διαμόρφωση της μεταγενέστερης γονιδιακής έκφρασης.

Επιπλέον, το p53 είναι ένα καλά μελετημένο γονίδιο καταστολής του καρκίνου. Η πρωτεΐνη που κωδικοποιείται από το p53 έχει ρόλο καταστολής του καρκίνου υπό κανονικές συνθήκες και προάγοντας την ανάπτυξη του καρκίνου όταν εμφανίζονται μεταλλάξεις. Όταν συμβαίνει βλάβη στο DNA, το p53 ξεκινά την απόπτωση των κυττάρων ρυθμίζοντας την έκφραση γονιδίων. Όταν συμβαίνει βλάβη του DNA, η χρωματίνη μετατρέπεται από απρόσιτη σε προσβάσιμη κατάσταση και ταυτόχρονα ενεργοποιείται το γονίδιο p53 για τη διατήρηση της σταθερότητας του γονιδιώματος.

ΚΕΦΑΛΑΙΟ 5: ΣΥΝΔΥΑΣΜΟΣ ΤΕΧΝΟΛΟΓΙΩΝ

5.1. Σημασία συνδυασμού τεχνολογιών

Η ενσωμάτωση των τεχνολογιών στην ανάλυση single – cell είναι ζωτικής σημασίας για διάφορους λόγους, καθένας από τους οποίους συμβάλλει σε μια βαθύτερη και πιο ολοκληρωμένη κατανόηση της κυτταρικής βιολογίας. Πρώτον, ενισχύει την ανάλυση και την ακρίβεια, επιτρέποντας την εξέταση μεμονωμένων κυττάρων και προσφέροντας γνώσεις για την κυτταρική ετερογένεια που δεν μπορεί να προσφέρει η bulk ανάλυση. Με τη χρήση συνδυασμού τεχνολογιών, τα δεδομένα συλλέγονται με υψηλή ανάλυση και ακρίβεια. Δεύτερον, η ενσωμάτωση αυτών των τεχνολογιών διευκολύνει την ολοκληρωμένη συλλογή δεδομένων πολλαπλών ομικών. Αυτή η προσέγγιση περιλαμβάνει τη γονιδιωματική, τη μεταγραφική, την πρωτεομική και τη μεταβολομική σε επίπεδο ενός κυττάρου, επιτρέποντας τη συσχέτιση διαφορετικών τύπων μοριακών δεδομένων, και οδηγώντας σε πιο ισχυρά και ενημερωτικά συμπεράσματα.

Η κατανόηση της κυτταρικής ετερογένειας είναι ένα άλλο σημαντικό όφελος. Οι κυτταρικοί πληθυσμοί συχνά αποτελούνται από ετερογενείς τύπους κυττάρων με ξεχωριστές λειτουργίες και καταστάσεις. Η ενσωμάτωση πολλαπλών τεχνολογιών χαρακτηρίζει με ακρίβεια αυτούς τους υποπληθυσμούς, διευκρινίζοντας τους ρόλους τους στην υγεία και τις ασθένειες και προσδιορίζοντας σπάνιους τύπους κυττάρων ή καταστάσεις ζωτικής σημασίας για συγκεκριμένες βιολογικές διεργασίες. Επιπλέον, οι συνδυαστικές προσεγγίσεις παρέχουν γνώσεις και λειτουργική ανάλυση συνδέοντας μοριακά προφίλ με κυτταρικούς φαινότυπους. Για παράδειγμα, ο συνδυασμός της αλληλουχίας ενός κυττάρου με την απεικόνιση ζωντανών κυττάρων συσχετίζει τα πρότυπα έκφρασης γονιδίων με δυναμικές κυτταρικές συμπεριφορές, προσφέροντας βαθύτερες γνώσεις για τις κυτταρικές λειτουργίες.

Αυτή η ενσωμάτωση οδηγεί επίσης σε βελτιωμένες διαγνωστικές και θεραπευτικές στρατηγικές. Σε ασθένειες όπως ο καρκίνος, όπου η ετερογένεια του όγκου επηρεάζει την αντοχή στη θεραπεία και την εξέλιξη της νόσου, οι ολοκληρωμένες τεχνολογίες μονοκυττάρων μπορούν να

εντοπίσουν βιοδείκτες και θεραπευτικούς στόχους, επιτρέποντας πιο ακριβείς και εξατομικευμένες προσεγγίσεις αντιμετώπισης. Επιπλέον, η ενσωμάτωση τεχνολογιών οδηγεί στην ανάπτυξη νέων αναλυτικών εργαλείων και υπολογιστικών μεθόδων που έχουν σχεδιαστεί για να χειρίζονται και να ερμηνεύουν σύνθετα σύνολα δεδομένων και βελτιώνουν την ακρίβεια των αναλύσεων και ενισχύουν την αναπαραγωγιμότητα και την επεκτασιμότητα των πειραμάτων.

Η χρήση πολλαπλών τεχνολογιών για τη μελέτη των ίδιων κυττάρων ή κυτταρικών πληθυσμών παρέχει ένα μέσο επικύρωσης δεδομένων και διασταυρούμενης επαλήθευσης και με τον τρόπο αυτό ενισχύεται η εμπιστοσύνη στα αποτελέσματα. Τα ευρήματα από την αλληλούχιση RNA ενός κυττάρου μπορούν να επικυρωθούν μέσω πρωτεομικών ή δεδομένων απεικόνισης, εξασφαλίζοντας ότι οι παρατηρήσεις δεν είναι αντικείμενα μιας μόνο μεθόδου.

Τέλος, η ενσωμάτωση διαφορετικών τεχνολογιών ανοίγει νέους δρόμους για την εξερεύνηση σύνθετων βιολογικών ζητημάτων που προηγουμένως δεν ήταν προσβάσιμα. Αυτό περιλαμβάνει τη μελέτη της χωρικής οργάνωσης των κυττάρων μέσα στους ιστούς, την κατανόηση των κυτταρικών αντιδράσεων στις περιβαλλοντικές αλλαγές και την αποκάλυψη της μοριακής βάσης της κυτταρικής διαφοροποίησης και ανάπτυξης.

5.2. Παραδείγματα συνδυαστικής χρήσης τεχνολογιών

5.2.1. Συνδυαστική χρήση scRNA-seq και ATAC-seq στην περίπτωση της παιδικής λευχαιμίας

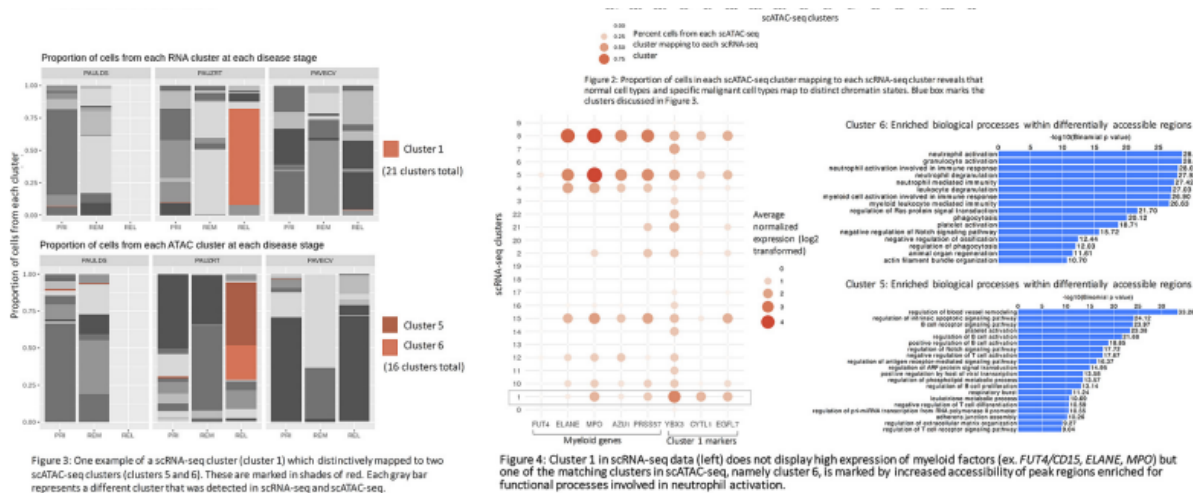
Η παιδική λευχαιμία είναι μια ετερογενής ασθένεια, στην οποία η αντίσταση στη θεραπεία παραμένει ένα άλυτο πρόβλημα που είναι υπεύθυνο για τους περισσότερους θανάτους. Πρόσφατα έγινε αντιληπτό ότι η αντίσταση μπορεί να καθοδηγείται από μηχανισμούς που εκτείνονται πέρα από τις μεταλλάξεις του DNA. Οι μεταγραφικές μεταβολές εντός συγκεκριμένων πρωτόγονων και δεσμευμένων κυτταρικών τύπων σε όγκους, οι οποίες μπορεί να συνοδεύονται από μεταβολές στη δομή και την τοπολογία της χρωματίνης, μπορούν επίσης να συμβάλουν στην εξέλιξη της νόσου. Για να μελετηθούν [39] τέτοιες αλλαγές σε επίπεδο μονοκυττάρων, αναλύθηκαν scRNA-seq δεδομένα και αντιστοιχήθηκαν δεδομένα scATAC-seq

από δείγματα πρωτοπαθούς, ύφεσης και/ή υποτροπής που ελήφθησαν από τρεις ασθενείς, οι οποίοι εντάχθηκαν σε κλινικές δοκιμές. Χρησιμοποιώντας τις πλατφόρμες γονιδιωματικής 10x, σχεδιάστηκαν συνολικά 39.738 κύτταρα χρησιμοποιώντας scRNA-seq (~4.826 κύτταρα ανά δείγμα, 1.571 γονίδια ανά κύτταρο) και 46.580 κύτταρα και 197.128 peaks χρησιμοποιώντας scATAC-seq (~6.718 κύτταρα ανά δείγμα, 5.628 μοναδικές αναγνώσεις ανά κύτταρο). Στη συνέχεια ενσωματώθηκαν αυτοί οι τύποι δεδομένων για να προσδιοριστεί ο βαθμός στον οποίο αυτές οι δύο μέθοδοι επιβεβαιώθηκαν ή/και συμπληρώθηκαν μεταξύ τους σε αναλύσεις αυτών των ληφθέντων δειγμάτων.

Οι υποπληθυσμοί κυττάρων που ανιχνεύθηκαν στην scRNA-seq μέσω της ομαδοποίησης Leiden σε ένα γράφημα ήταν γενικά συνεπείς με πρόσφατες παρατηρήσεις κακοήθων και φυσιολογικών τύπων κυττάρων που ανιχνεύθηκαν σε τμήματα μυελού των οστών και περιφερικού αίματος. Οι κακοήθεις υποπληθυσμοί σε πρωτογενή και υποτροπιάζοντα στάδια παρουσίασαν παρόμοια επίπεδα ποικιλίας κυτταρικού τύπου κατά μήκος της μυελοειδούς γενεαλογίας. Αυτά περιλάμβαναν αιμοποιητικά βλαστοειδή κύτταρα, προγονικούς, προγονικούς κοκκιοκυττάρων-μονοκυττάρων, μονοκύτταρα και υποπληθυσμούς που μοιάζουν με δενδριτικά κύτταρα. Τα δείγματα ύφεσης φάνηκαν να περιέχουν φυσιολογικούς τύπους αιμοσφαιρίων, συμπεριλαμβανομένων των αιμοπεταλίων και των ερυθροκυττάρων. Ωστόσο, παρατηρήθηκαν επίσης υποτιθέμενοι κακοήθεις συμβατικοί υποπληθυσμοί δενδριτικών κυττάρων σε ύφεση (50% και 16% στα αντίστοιχα δείγματα), σημειώνοντας ότι αυτά τα κύτταρα εμφάνισαν αυξημένη έκφραση γονιδίων που εμπλέκονται στην παρουσίαση αντιγόνου και την επεξεργασία λυσοσωμικών πρωτεϊνών.

Για την ενσωμάτωση των scATAC-seq δεδομένων με τα δεδομένα scRNA-seq πραγματοποιήθηκε ομαδοποίηση μετασχηματισμένων και μειωμένων δεδομένων scATAC-seq μέσω επαναληπτικής λανθάνουσας σημασιολογικής ευρετηρίασης, και ευθυγράμμιση, χρησιμοποιώντας ανάλυση συσχέτισης. Οι υποπληθυσμοί scRNA-seq που κυριαρχούνται από κακοήθη κύτταρα έδειξαν μεταβλητότητα στη χαρτογράφηση σε διακριτές καταστάσεις χρωματίνης, με μερικές αξιοσημείωτες εξαιρέσεις. Μία τέτοια εξαίρεση είναι ένας υποπληθυσμός στην scRNA-seq, που βρέθηκε κυρίως σε υποτροπή, που χαρακτηρίζεται από υψηλή έκφραση γονιδίων που εμπλέκονται σε κυτταρικές διεργασίες πολλαπλασιασμού και αυξητικού παράγοντα όπως ο YBX3 (συνδέεται με τον υποκινητή GM-CSF), ο CYTL1 και ο

EGFL7 (ρυθμιστής αγγειογένεσης). Ο υποπληθυσμός scRNA-seq, ωστόσο, δεν εμφάνισε υψηλή έκφραση μυελοειδών/κοκκιοκυττάρων παραγόντων όπως CD15, ELANE και MPO, ίσως σύμφωνη με την ιδέα ότι τέτοια μεταγραφικά προγράμματα μπορεί να προετοιμαστούν αλλά δεν έχουν ακόμη ενεργοποιηθεί μέσα σε αυτά τα κακοήγη κύτταρα.



Εικόνα 5: Παρουσίαση αποτελεσμάτων

5.2.2. Ολοκληρωμένη μονοκυτταρική ανάλυση RNA-Seq και ATAC-Seq επιθηλιακών κυττάρων κερατοειδούς ποντικού

Η ομοιότητα του επιθηλίου του κερατοειδούς διατηρείται με συντονισμένη γονιδιακή έκφραση σε διαφορετικούς κυτταρικούς πληθυσμούς, αλλά τα ρυθμιστικά προγράμματα γονιδίων που αποτελούν τη βάση αυτής της κυτταρικής ποικιλομορφίας παραμένουν να χαρακτηριστούν. Στη περίπτωση αυτής της μελέτης [40], εφαρμόστηκε ανάλυση πολλαπλών μονοκυττάρων για να οριοθετηθεί το γονιδιακό ρυθμιστικό προφίλ των επιθηλιακών κυττάρων του κερατοειδούς ποντικού υπό κανονική ομοιότητα.

Όσο αφορά τη μεθοδολογία που ακολουθήθηκε, μεμονωμένα κύτταρα που απομονώθηκαν από το επιθήλιο του κερατοειδούς (με οριακό επιπεφυκότα) ενήλικων ποντικών υποβλήθηκαν σε scRNA-seq και scATAC-seq ανάλυση χρησιμοποιώντας την πλατφόρμα γονιδιωματικής 10x. Οι τύποι κυττάρων ομαδοποιήθηκαν με τη μέθοδο απεικόνιση που βασίζεται στη χρήση

γραφήματος. Τα σύνολα δεδομένων scRNA-seq και scATAC-seq ενσωματώθηκαν μετά τον αγωγό ολοκλήρωσης που περιγράφεται στα ArchR και Seurat.

Όσο αφορά τα αποτελέσματα, χαρακτηρίστηκαν διαφορετικοί τύποι επιθηλιακών κυττάρων κερατοειδούς με βάση τις υπογραφές γονιδιακής έκφρασης και την προσβασιμότητα της χρωματίνης. Διαπιστώθηκε ότι οι περιοχές προσβασιμότητας ειδικού τύπου κυττάρων εντοπίστηκαν κυρίως σε απομακρυσμένες περιοχές, υποδηλώνοντας βασικούς ρόλους απομακρυσμένων ρυθμιστικών στοιχείων στον προσδιορισμό της ποικιλομορφίας των επιθηλιακών κυττάρων του κερατοειδούς. Οι αναλύσεις τροχιάς αποκάλυψαν ένα συνεχές της μετάβασης της κυτταρικής κατάστασης και υψηλότερο συντονισμό μεταξύ της προσβασιμότητας του μοτίβου του παράγοντα μεταγραφής (TF) και της γονιδιακής έκφρασης κατά τη διάρκεια της διαφοροποίησης των επιθηλιακών κυττάρων του κερατοειδούς.

5.2.3. Συνδυαστική χρήση scRNA-Seq και ATAC-Seq σε ασθενείς με αγκυλοποιητική σπονδυλίτιδα

Ερευνητικές μελέτες για την αγκυλοποιητική σπονδυλίτιδα (AS) έχουν εντοπίσει περισσότερα από 100 παθογόνα γονίδια. Η οικοδόμηση μιας γέφυρας μεταξύ αυτών των γονιδίων και των βιολογικά στοχευμένων θεραπειών είναι το τρέχον ερευνητικό σημείο. Για τον σκοπό αυτό [41], ενσωματώθηκε μια αλληλουχία χρωματίνης προσπελάσιμης με τρανσποζάση (scATAC-seq) και μια αλληλουχία RNA ενός κυττάρου (scRNA-seq) προκειμένου να διερευνηθούν τα βασικά γονίδια και οι σχετικοί μηχανισμοί που σχετίζονται με την παθογένεια AS.

Ως αποτέλεσμα αυτής της πειραματικής προσπάθειας, εντοπίστηκαν 18 κυτταρικοί τύποι σε περιφερικά μονοπύρρηνα κύτταρα από ασθενείς με AS και φυσιολογικούς ανθρώπους και συνοψίστηκαν τα μη φυσιολογικά γονίδια ειδικού τύπου κυττάρου με scRNA-seq. Είναι ενδιαφέρον ότι διαπιστώθηκε ότι το παθογόνο γονίδιο NFκB που εμπλέκεται στην εξέλιξη AS προήλθε από CD8+ T κύτταρα. Επιπλέον, τα αποτελέσματα scATAC-seq επιβεβαίωσαν τις μη φυσιολογικές προσβάσιμες θέσεις σύνδεσης των μεταγραφικών παραγόντων FOS, JUN και JUNB. Η τελική διαλογή μαγνητικών σφαιριδίων και η ποσοτική PCR σε πραγματικό χρόνο (RT-

qPCR) επιβεβαίωσαν ότι τα NFκB, FOS, JUN και JUNB σε CD8+ T κύτταρα διέφεραν στην ομάδα AS.

5.3. Προκλήσεις

Η ενσωμάτωση των τεχνολογιών στην single - cell ανάλυση, παρά τα πολυάριθμα πλεονεκτήματά της, παρουσιάζει αρκετές σημαντικές προκλήσεις. Μία από τις κύριες προκλήσεις είναι η τεχνική πολυπλοκότητα που συνεπάγεται ο συγχρονισμός πολλαπλών μεθόδων. Κάθε τεχνολογία έχει τις ειδικές απαιτήσεις της για την προετοιμασία δειγμάτων, την απόκτηση δεδομένων και την επεξεργασία. Η διασφάλιση της συμβατότητας και της απρόσκοπτης ολοκλήρωσης μεταξύ αυτών των τεχνολογιών μπορεί να είναι τεχνικά απαιτητική και δαπανηρή όσο αφορά τη χρήση πόρων. Επιπλέον, η single – cell ανάλυση συχνά απαιτεί χειρισμό μικροσκοπικών ποσοτήτων βιολογικού υλικού, καθιστώντας τη επιρρεπή σε απώλεια απώλεια δειγμάτων, γεγονός που μπορεί να θέσει σε κίνδυνο την ποιότητα και την αναπαραγωγιμότητα των δεδομένων.

Επιπλέον, η τεράστια ποσότητα δεδομένων που παράγονται από τεχνολογίες τέτοιου τύπου απαιτεί προηγμένα υπολογιστικά εργαλεία και σημαντική υπολογιστική ισχύ για επεξεργασία, αποθήκευση και ανάλυση. Η ανάπτυξη ισχυρών αλγορίθμων που μπορούν να ενσωματώσουν και να ερμηνεύσουν με ακρίβεια δεδομένα πολλαπλών μεθόδων είναι ένα πολύπλοκο έργο. Επιπλέον, η ερμηνεία των συνδυαστικών δεδομένων απαιτεί εμπειρογνωμοσύνη σε διάφορους τομείς, συμπεριλαμβανομένης της γονιδιωματικής, της βιοπληροφορικής και της βιολογίας συστημάτων, γεγονός που μπορεί να αποτελέσει εμπόδιο για πολλές ερευνητικές ομάδες. Η ενσωμάτωση δεδομένων εγείρει επίσης ζητήματα που σχετίζονται με την τυποποίηση και την ομαλοποίηση, καθώς διαφορετικές τεχνολογίες μπορούν να παράγουν δεδομένα σε διαφορετικές κλίμακες και μορφές, περιπλέκοντας τις διαδικασίες σύγκρισης και ολοκλήρωσης. Η υπέρβαση αυτών των προκλήσεων είναι απαραίτητη για την πλήρη αξιοποίηση των δυνατοτήτων της ολοκληρωμένης single – cell ανάλυσης στην προώθηση της βιολογικής έρευνας και των ιατρικών εφαρμογών.

ΚΕΦΑΛΑΙΟ 6 : ΣΥΜΠΕΡΑΣΜΑΤΑ

Στα πλαίσια της συγκεκριμένης εργασίας, πραγματοποιήθηκε εκτενής βιβλιογραφική ανασκόπηση όσο αφορά τεχνολογίες single – cell, δίνοντας έμφαση στα χαρακτηριστικά τους και στις μεθοδολογίες που χρησιμοποιούν. Αυτό που διαπιστώθηκε αρχικά είναι ότι έχει σημειωθεί σημαντική πρόοδος, συμβάλλοντας με τον τρόπο αυτό στην εξέλιξη της της ιατρικής και βιολογικής έρευνας.

Το μέλλον των τεχνολογιών αυτών αναμένεται να έχει σημαντικές επιπτώσεις στην κατανόηση της βιολογίας και της ιατρικής, με αρκετές βασικές εξελίξεις στον ορίζοντα. Ένας σημαντικός τομέας ανάπτυξης είναι η ενσωμάτωση των πολύ-ομικών προσεγγίσεων σε επίπεδο ενός κυττάρου. Συνδυάζοντας δεδομένα από τη γονιδιωματική, τη μεταγραφική, την πρωτεομική και τη μεταβολομική, οι ερευνητές μπορούν να αποκτήσουν μια ολοκληρωμένη εικόνα των κυτταρικών καταστάσεων και λειτουργιών. Αυτή η ολιστική προσέγγιση θα διευκολύνει την ανακάλυψη νέων βιοδεικτών και θεραπευτικών προσεγγίσεων, ενισχύοντας εξατομικευμένες στρατηγικές ιατρικής. Επιπλέον, οι εξελίξεις στα υπολογιστικά εργαλεία και την τεχνητή νοημοσύνη αναμένεται να βελτιώσουν την ανάλυση και την ερμηνεία σύνθετων δεδομένων ενός κυττάρου, καθιστώντας δυνατή την αποκάλυψη των περίπλοκων ρυθμιστικών δικτύων που διέπουν τη συμπεριφορά των κυττάρων.

Μια άλλη πολλά υποσχόμενη κατεύθυνση είναι η ενίσχυση των χωρικά προσανατολισμένων τεχνολογιών. Οι τρέχουσες μέθοδοι όπως η χωρική μεταγραφική παρέχουν ήδη πληροφορίες για την χωρική οργάνωση των κυττάρων μέσα στους ιστούς. Οι μελλοντικές καινοτομίες πιθανότατα θα προσφέρουν ακόμη υψηλότερες δυνατότητες ανάλυσης, επιτρέποντας στους ερευνητές να χαρτογραφήσουν τις κυτταρικές αλληλεπιδράσεις και τα μικροπεριβάλλοντα με υψηλού επιπέδου λεπτομέρειες. Αυτές οι χωρικά προσανατολισμένες τεχνικές θα είναι κρίσιμης σημασίας για την κατανόηση της αρχιτεκτονικής των ιστών στην υγεία και τις ασθένειες, ενδεχομένως οδηγώντας σε νέα διαγνωστικά εργαλεία και θεραπευτικές προσεγγίσεις. Επιπλέον, η ενσωμάτωση χωρικών δεδομένων με απεικόνιση χρονικού διαστήματος και παρακολούθηση ζωντανών κυττάρων θα μπορούσε να παρέχει δυναμικές προσεγγίσεις των κυτταρικών διεργασιών, προσφέροντας πληροφορίες για τη μετανάστευση των κυττάρων, τη διαφοροποίηση και την ανάπτυξη ιστών σε πραγματικό χρόνο.

Η εφαρμογή single – cell τεχνολογιών σε κλινικά περιβάλλοντα αναμένεται επίσης να επεκταθεί σημαντικά. Η αλληλούχιση ενός κυττάρου θα μπορούσε να γίνει ένα συνηθισμένο τμήμα των διαγνωστικών διαδικασιών, ιδιαίτερα στην ογκολογία, όπου μπορεί να βοηθήσει στην έγκαιρη ανίχνευση του καρκίνου, στην παρακολούθηση της ελάχιστης υπολειμματικής νόσου και στην κατανόηση της ετερογένειας του όγκου. Επιπλέον, οι αναλύσεις αυτές μπορούν να συμβάλουν στην ανάπτυξη πιο αποτελεσματικών ανοσοθεραπειών με τον εντοπισμό συγκεκριμένων υποσυγκροτημάτων ανοσοκυττάρων που εμπλέκονται στην εμφάνιση ασθενειών. Καθώς οι τεχνολογίες γίνονται πιο προσιτές και επεκτάσιμες, η χρήση τους σε πληθυσμιακές μελέτες μεγάλης κλίμακας θα παράσχει πολύτιμα δεδομένα σχετικά με την ανθρώπινη ποικιλομορφία.

Συμπερασματικά, η συνεχής εξέλιξη των τεχνολογιών single – cell έχει τεράστιες δυνατότητες για τη βιολογική έρευνα και την κλινική πρακτική, προσφέροντας βαθύτερες γνώσεις σχετικά με τις θεμελιώδεις μονάδες της ζωής.

ΒΙΒΛΙΟΓΡΑΦΙΑ – ΠΗΓΕΣ

- [1]. Carangelo, G., Magi, A., & Semeraro, R. (2022). From multitude to singularity: An up-to-date overview of scRNA-seq data generation and analysis. *Frontiers in Genetics*, 13, 994069.
- [2]. Cock, P. J., Fields, C. J., Goto, N., Heuer, M. L., & Rice, P. M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic acids research*, 38(6), 1767-1771.
- [3]. Kivioja, T., Vähärautio, A., Karlsson, K., Bonke, M., Enge, M., Linnarsson, S., & Taipale, J. (2012). Counting absolute numbers of molecules using unique molecular identifiers. *Nature methods*, 9(1), 72-74.
- [4]. Bushel, P. R., Ferguson, S. S., Ramaiahgari, S. C., Paules, R. S., & Auerbach, S. S. (2020). Comparison of normalization methods for analysis of TempO-Seq Targeted RNA sequencing data. *Frontiers in Genetics*, 11, 594.
- [5]. Erlich, H. A. (1989). *PCR technology* (Vol. 246). New York: Stockton press.
- [6]. Kujawa, T., Marczyk, M., & Polanska, J. (2022). Influence of single-cell RNA sequencing data integration on the performance of differential gene expression analysis. *Frontiers in Genetics*, 13, 1009316.
- [7]. Tian, T. (2019). *Model-Based Deep Autoencoders for Characterizing Discrete Data with Application to Genomic Data Analysis* (Doctoral dissertation, New Jersey Institute of Technology).
- [8]. Kasianchuk, N., Kukuruza, Y., Ostash, V., Boshtova, A., & Mykhailichenko, D. (2023, November). Scrutinised and Compared: HVG Identification Methods in Terms of Common Metrics. In *Proceedings of International Conference on Applied Innovation in IT* (Vol. 11, No. 2, pp. 59-65). Anhalt University of Applied Sciences.
- [9]. Wolf, F. A., Angerer, P., & Theis, F. J. (2018). SCANPY: large-scale single-cell gene expression data analysis. *Genome biology*, 19, 1-5.
- [10]. Liu, Z. (2020). Visualizing single-cell RNA-seq data with semisupervised principal component analysis. *International journal of molecular sciences*, 21(16), 5797.

- [11]. Linderman, G. C., Rachh, M., Hoskins, J. G., Steinerberger, S., & Kluger, Y. (2019). Fast interpolation-based t-SNE for improved visualization of single-cell RNA-seq data. *Nature methods*, 16(3), 243-245.
- [12]. Becht, E., McInnes, L., Healy, J., Dutertre, C. A., Kwok, I. W., Ng, L. G., ... & Newell, E. W. (2019). Dimensionality reduction for visualizing single-cell data using UMAP. *Nature biotechnology*, 37(1), 38-44.
- [13]. Street, K., Risso, D., Fletcher, R. B., Das, D., Ngai, J., Yosef, N., ... & Dudoit, S. (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC genomics*, 19, 1-16.
- [14]. Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H. A., & Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. *Nature methods*, 14(10), 979-982.
- [15]. Wolf, F. A., Hamey, F. K., Plass, M., Solana, J., Dahlin, J. S., Göttgens, B., ... & Theis, F. J. (2019). PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome biology*, 20, 1-9.
- [16]. Shi, J., & Walker, M. G. (2007). Gene set enrichment analysis (GSEA) for interpreting gene expression profiles. *Current Bioinformatics*, 2(2), 133-137.
- [17]. Zou, J., Deng, F., Wang, M., Zhang, Z., Liu, Z., Zhang, X., ... & Hao, J. (2022). scCODE: an R package for data-specific differentially expressed gene detection on single-cell RNA-sequencing data. *Briefings in Bioinformatics*, 23(5), bbac180.
- [18]. Liu, S., Wang, Z., Zhu, R., Wang, F., Cheng, Y., & Liu, Y. (2021). Three differential expression analysis methods for RNA sequencing: limma, EdgeR, DESeq2. *JoVE (Journal of Visualized Experiments)*, (175), e62528.
- [19]. Henrichsen, C. N., Chaignat, E., & Reymond, A. (2009). Copy number variants, diseases and gene expression. *Human molecular genetics*, 18(R1), R1-R8.
- [20]. Replogle, J. M., Norman, T. M., Xu, A., Hussmann, J. A., Chen, J., Cogan, J. Z., ... & Adamson, B. (2020). Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nature biotechnology*, 38(8), 954-961.
- [21]. Williams, C. G., Lee, H. J., Asatsuma, T., Vento-Tormo, R., & Haque, A. (2022). An introduction to spatial transcriptomics for biomedical research. *Genome Medicine*, 14(1), 68.

- [22]. Rao, N., Clark, S., & Habern, O. (2020). Bridging genomics and tissue pathology: 10x genomics explores new frontiers with the visium spatial gene expression solution. *Genetic Engineering & Biotechnology News*, 40(2), 50-51.
- [23]. Ke, R., Mignardi, M., Pacureanu, A., Svedlund, J., Botling, J., Wählby, C., & Nilsson, M. (2013). In situ sequencing for RNA analysis in preserved tissue and cells. *Nature methods*, 10(9), 857-860.
- [24]. Cassidy, A., & Jones, J. (2014). Developments in in situ hybridisation. *Methods*, 70(1), 39-45.
- [25]. Tsanov, N., Samacoits, A., Chouaib, R., Traboulsi, A. M., Gostan, T., Weber, C., ... & Mueller, F. (2016). smiFISH and FISH-quant—a flexible single RNA detection approach with super-resolution capability. *Nucleic acids research*, 44(22), e165-e165.
- [26]. Yan, Y., Yang, L., Meng, L., Su, H., Zhou, C., Yu, L., ... & Gao, J. (2023). Introduction to bioimaging-based spatial multi-omic novel methods. *Quantitative Biology*, 11(3), 231-245.
- [27]. Kim, T., Lim, C. S., & Kaang, B. K. (2015). Cell type-specific gene expression profiling in brain tissue: comparison between TRAP, LCM and RNA-seq. *BMB reports*, 48(7), 388.
- [28]. Chandramohan, L., Au, Q., Nagy, M., Parnell, E., & Peters, T. (2021). comprehensive analysis of immuno oncology markers in the tumor microenvironment of solid tumor samples using GeoMxTM digital spatial profiler (DSP) and MultiOmyxTM hyperplexed immunofluorescence (IF). *Cancer Research*, 81(13_Supplement), 2771-2771.
- [29]. Zhao, E., Stone, M. R., Ren, X., Guenthoer, J., Smythe, K. S., Pulliam, T., ... & Gottardo, R. (2021). Spatial transcriptomics at subspot resolution with BayesSpace. *Nature biotechnology*, 39(11), 1375-1384.
- [30]. Bergenstråhle, L., He, B., Bergenstråhle, J., Andersson, A., Lundeberg, J., Zou, J., & Maaskola, J. (2020). Super-resolved spatial transcriptomics by deep data fusion. *BioRxiv*, 2020-02.
- [31]. Elosua-Bayes, M., Nieto, P., Mereu, E., Gut, I., & Heyn, H. (2021). SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic acids research*, 49(9), e50-e50.
- [32]. Cang, Z., & Nie, Q. (2020). Inferring spatial and signaling relationships between cells from single cell transcriptomic data. *Nature communications*, 11(1), 2084.

- [33]. Li, H., Zhou, J., Li, Z., Chen, S., Liao, X., Zhang, B., ... & Gao, X. (2023). A comprehensive benchmarking with practical guidelines for cellular deconvolution of spatial transcriptomics. *Nature Communications*, 14(1), 1548.
- [34]. Cabello-Aguilar, S., Alame, M., Kon-Sun-Tack, F., Fau, C., Lacroix, M., & Colinge, J. (2020). SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. *Nucleic acids research*, 48(10), e55-e55.
- [35]. Yan, F., Powell, D. R., Curtis, D. J., & Wong, N. C. (2020). From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. *Genome biology*, 21, 1-16.
- [36]. Sun, Y., Miao, N., & Sun, T. (2019). Detect accessible chromatin using ATAC-sequencing, from principle to applications. *Hereditas*, 156, 1-9.
- [37]. Chen, H., Lareau, C., Andreani, T., Vinyard, M. E., Garcia, S. P., Clement, K., ... & Pinello, L. (2019). Assessment of computational methods for the analysis of single-cell ATAC-seq data. *Genome biology*, 20, 1-25.
- [38]. Spektor, R., Tippens, N. D., Mimoso, C. A., & Soloway, P. D. (2019). methyl-ATAC-seq measures DNA methylation at accessible chromatin. *Genome research*, 29(6), 969-977.
- [39]. Wei, L., Trinh, D., Ries, R. E., Jin, D., Corbett, R. D., Smith, J. L., ... & Marra, M. A. (2020). Integrative Analysis of Single-Cell RNA-Seq and ATAC-Seq Data across Treatment Time Points in Pediatric AML. *Blood*, 136, 29.
- [40]. Lu, Z. J., Ye, J. G., Wang, D. L., Li, M. K., Zhang, Q. K., Liu, Z., ... & Zheng, Y. F. (2023). Integrative single-cell RNA-Seq and ATAC-seq analysis of mouse corneal epithelial cells. *Investigative Ophthalmology & Visual Science*, 64(3), 30-30.
- [41]. Xu, H., Yu, H., Liu, L., Wu, H., Zhang, C., Cai, W., ... & Dai, Y. (2021). Integrative single-cell RNA-seq and ATAC-seq analysis of peripheral mononuclear cells in patients with ankylosing spondylitis. *Frontiers in Immunology*, 12, 760381.